Field Crops Research 228 (2018) 1–19

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

The response of process-based agro-ecosystem models to within-field variability in site conditions

Evelyn Wallor[a,*], Kurt-Christian Kersebaum[a], Domenico Ventrella[b], Marco Bindi[c], Davide Cammarano[d], Elsa Coucheney[e], Thomas Gaiser[f], Pasquale Garofalo[b], Luisa Giglio[b], Pietro Giola[g], Munir P. Hoffmann[h], Ileana Iocola[g], Marcos Lana[a], Elisabet Lewan[e], Ganga Ram Maharjan[f], Marco Moriondo[i], Laura Mula[g], Claas Nendel[a], Eva Pohankova[j,k], Pier Paolo Roggero[g], Mirek Trnka[j,k], Giacomo Trombi[c]

[a] *Leibniz Centre for Agricultural Landscape Research (ZALF), Institute of Landscape Systems Analysis, Eberswalder Str. 84, 15374 Müncheberg, Germany*
[b] *Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Centro di Ricerca Agricoltura e Ambiente (CREA-AA), Via Celso Ulpiani 5, 70125 Bari, Italy*
[c] *University of Florence, Department of Agri-Food Production and Environmental Sciences (DISPAA), 50144 Firenze, Italy*
[d] *The James Hutton Institute, Invergowrie Dundee DD2 5DA, United Kingdom*
[e] *Swedish University of Agricultural Science (SLU), Department of Soil and Environment, Box 7014, 75007 Uppsala, Sweden*
[f] *University of Bonn, Institute of Crop Science and Resource Conservation Crop Science Group, Katzenburgweg 5, 53115 Bonn, Germany*
[g] *Università degli studi di Sassari (UNISS), Nucleo di Ricera sulla Desertificazione e Dipartimento di Agraria, Viale Italia 39, 07100 Sassari, Italy*
[h] *Georg-August-University Göttingen, Tropical Plant Production and Agricultural Systems Modelling, 37077 Göttingen, Germany*
[i] *National Research Council, Institute of Biometeorology of the National Research Council (CNR-IBIMET), 50145 Firenze, Italy*
[j] *Global Change Research Institute of the Czech Academy of Sciences, Belidla 986/4a, 603 00 Brno, Czech Republic*
[k] *Mendel University, Institute of Agriculture Systems and Bioclimatology, Zemedelska 1, 61300 Brno, Czech Republic*

A R T I C L E I N F O

Keywords:
Spatial variability
Soil
Crop model
Model sensitivity

A B S T R A C T

Process-oriented agro-ecosystem models are increasingly applied to assess crop management options or impacts of climate change on agricultural production, food security and ecosystem services. Thereby, the aggregation of initial soil and climate information is a widely used approach for performing simulations at larger scales such as regions, nations or even globally. In this context, the ability of models to respond to different site conditions is essential for high quality impact assessment through the use of modelling tools. As part of a model inter-comparison the present study investigated models' yield response on variable site conditions using data sets from two well-documented fields, one located in Germany and one in Italy. The fields were sampled at 60 and 100 grid points, respectively, and soil and crop variables were recorded at varying intensity for the entire simulation period covering three growing seasons. The data was provided successively to the participating modelling groups in three calibration steps (a, b, and c) and the first growing season was considered for calibration. Model validation was based on these steps and each growing season as well as on the entire simulation period considering the soil state variables mineral nitrogen and water content (N, WC) as well as crop yield, biomass, and leaf area index (LAI). The WC was best depicted by the models, resulting in high correlation coefficients (r) up to 0.81 between simulated and observed values. The root mean square error (RMSE) of simulated N ranged from 20 kg ha^{-1} to 1072 kg ha^{-1} regarding all steps and growing seasons. The annual within-field variability of yields was better simulated by the models when observed subsoil information was provided. However, the RMSE ranged from 0.5 t ha^{-1} to 3.5 t ha^{-1} at the German field, and from 0.6 t ha^{-1} to 5.9 t ha^{-1} at the Italian field, respectively. It was found that intensified calibration did not necessarily lead to improved model output. Furthermore, single models showed specific inconsistencies in their algorithms when, for example, underestimated WC was associated with overestimated yields. In total, the sensitivity of models to spatially variable site conditions differed considerably. The importance of quality-assured soil and yield information for model improvement was highlighted.

* Corresponding author.
E-mail address: evelyn.wallor@zalf.de (E. Wallor).

https://doi.org/10.1016/j.fcr.2018.08.021
Received 15 December 2017; Received in revised form 15 August 2018; Accepted 26 August 2018
Available online 31 August 2018
0378-4290/ © 2018 Elsevier B.V. All rights reserved.

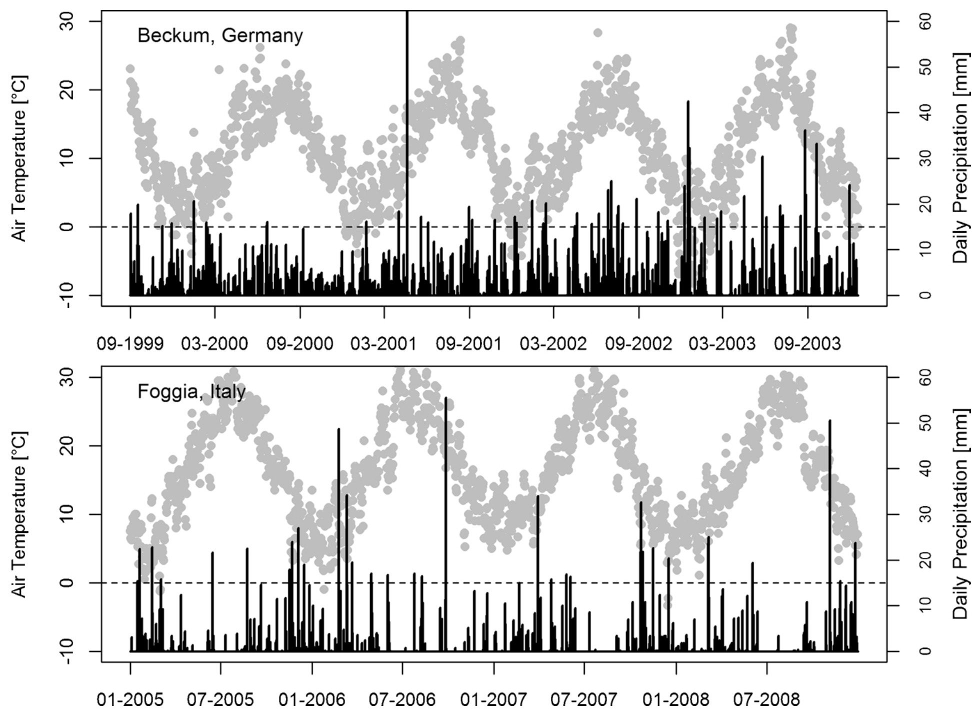

Fig. 1. Daily precipitation and averaged temperature of the fields Beckum (Germany; mean annual temperature = 11 °C, annual precipitation = 903 mm) and Foggia (Italy; 16 °C, 502 mm).

1. Introduction

Understanding the interaction between soil, crop and atmosphere is an essential prerequisite to describe the multiple functions of agro-ecosystems at different scales. Process-oriented, dynamic agro-ecosystem models are widely accepted tools to evaluate crop management options or to assess impacts of climate change on crop production, food security and ecosystem services (Ewert et al., 2015). Although mostly developed at field scale, crop growth models were increasingly applied for larger areas such as regions, nations, large watersheds and/or globally (Rosenzweig and Parry, 1994; Rötter and Van de Geijn, 1999; Easterling et al., 2007) to assess impacts and risks for crop production under different climate scenarios. Vulnerability of crop production to changing climate conditions is highly determined by the ability of the site to buffer periods of adverse climatic situations like water scarcity or excessive rainfall (Trnka et al., 2014). Therefore, impacts of climate change on crop production depend strongly on the site conditions and properties (Kersebaum and Nendel, 2014; Nendel et al., 2014). Moreover, seepage, nutrient losses and greenhouse gas emissions are affected, which might pollute natural resources and contribute to climate change. Therefore, the capability of models to reflect crop responses and water and nutrient dynamics under different site conditions is essential to properly assess climate impact on a regional scale. Spatial aggregation of soils can have a dominant influence of the outcome of regional yield simulations (Hoffmann et al., 2016). However, models showed significant differences in their response to soil and climate aggregation for crop yield (Hoffmann et al., 2016; Zhao et al., 2016), primary production (Kuhnert et al., 2017) as well as for soil organic carbon dynamics (Grosz et al., 2017). The variability of soil properties within a field often represents most of the variability of a whole landscape (De Wit and van Keulen, 1987) since processes of soil development vary according to almost every spatial and temporal scale. Hence, chemical, physical, and biological soil properties differ greatly even in their small-scale distribution (Pätzold et al., 2008; Selige et al., 2006; Kersebaum et al., 2002; Moore et al., 1993). With respect to agricultural land-use crop growth conditions are heterogeneously distributed in vertical and horizontal dimensions within fields influencing the local yield performance (Stafford et al., 1999; Pätzold et al., 2008; Gebbers and Adamchuk, 2010; Geesing et al., 2014). From environmental aspects, the water and nutrient holding capacity of individual soils vary. Hence, the risk of resource losses and contamination of adjacent ecosystems is site-specific (Kersebaum et al., 2005). This also implies that the inter-annual transfer of water and nutrients between crops in rotations and consequently the impact of the previous crop on water and nutrient supply differ. The aspect of crop rotations with consideration of carry-over effects have been rarely addressed in climate impact studies so far (Kollas et al., 2015; Yin et al., 2017).

To test the sensitivity of models to various site properties such as soil variability and hydrological boundary conditions, an ensemble of 11 different models were applied based on spatially variable data sets from precision farming of two sites in Europe covering a three year crop rotation. Simulations were performed using three levels of information for model calibration. The objective was a) to evaluate the sensitivity of the models in response to the variable soil information, and b) to analyse the consistency of their output among crop variables and between crop growth and soil water and nitrogen dynamics along the crop rotation comparing simulated model variables to spatial crop and soil observations.

2. Material and methods

2.1. Site description

Two sites with different soils and contrasting climatic conditions were selected for the study: a farmer´s field at Beckum/Germany with a 3 year wheat-wheat-triticale crop rotation and the field at Foggia/Italy cropped with durum wheat during three years. Daily weather was observed in close proximity (< 300 m) to both fields. Since both fields are in a flat landscape, we assume that weather variables can be assumed to be equal for the whole fields and surface run-off and water re-distribution between grid points are negligible. Daily records for precipitation and temperature during the period of investigation are provided in Fig. 1, management information is provided in Table 1.

Table 1
Management information of the fields Beckum and Foggia used for the simulation.

Date	Management
Beckum, Germany	
20.08.1999	harvest winter rape
17.09.1999	shallow tillage (8 cm depth, ploughing)
19.09.1999	sowing winter wheat (variety: Batis)
09.08.2000	harvest winter wheat
23.08.2000	seedbed preparation (16 cm depth, no ploughing)
24.09.2000	sowing winter wheat (variety: Batis)
23.07.2001	harvest winter wheat
14.08.2001	seedbed preparation (16 cm depth, no ploughing)
29.09.2001	sowing triticale (variety: Lamberto)
28.07.2002	harvest triticale
Foggia, Italy	
30.10.2005	ploughing (40 cm)
13.11.2005	disc-harrowing (15 cm)
07.12.2005	disc-harrowing (15 cm)
15.12.2005	sowing durum wheat (variety: Gargano)
26.06.2006	harvest durum wheat
28.10.2006	ploughing (40 cm)
07.11.2006	disc-harrowing (15 cm)
23.11.2006	disc-harrowing (15 cm)
26.11.2006	sowing durum wheat (variety: Gargano)
05.07.2007	harvest durum wheat
14.12.2007	direct sowing durum wheat (variety: Gargano)
05.07.2008	harvest durum wheat

2.1.1. *Beckum*

The experimental field is located in Beckum, North-Rhine Westphalia, Germany (51° 45′ N, 7° 59′ E, 98 m above sea level), and has a size of 20 ha. The elevation above sea level ranges from 96 to 102 m. The soil map shows a clear pattern of different soil types with large differences in texture ranging from silty sand to clay loam (Fig. 2). In the southern part of the field stone content increases in the lower part of the profile due to an ascending calcareous marl layer. Auger

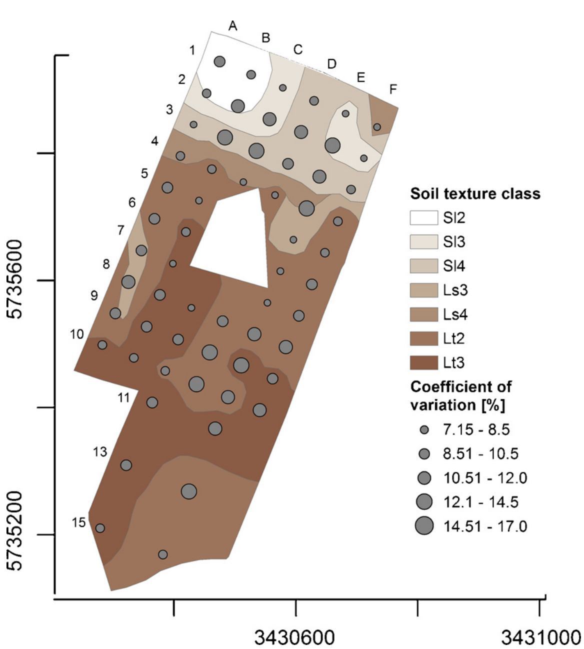

Fig. 2. Spatial distribution of soil texture at the field Beckum, Germany, averaged to a soil depth of 90 cm and grid points of the sampled grid (n = 60); Sl2 = slightly loamy sand, Sl3 = moderately loamy sand, Sl4 = strongly loamy sand, Ls3 = moderately sandy loam, Ls4 = strongly sandy loam, Lt2 = slightly clayey loam, Lt3 = moderately clayey loam; size of grid points represents the three-year average of the small scale coefficient of variation [%] of yields within the 25 diameter buffer zone around each grid point at the Beckum site.

sampling from 0 to 90 cm was performed in September 1999 after winter oilseed rape (*Brassica napus* L.) and before sowing of winter wheat (*Triticum aestivum* L.) at 60 locations in a 50·50 m, or at most in a 100·100 m grid, leaving out the forested central area of the field and the headland. From five locations within a radius of two meters around each grid point composite soil samples were formed and used to determine the spatial distribution of basic stable soil characteristics (texture, C_{org} and N_{org}) and to observe initial state variables (soil water and soil mineral nitrogen) to initiate the models. The samples were divided into three layers before soil analyses (0 to 30 cm, 30 to 60 cm, 60 to 90 cm), so that there were soil profile information for each of the 60 grid points. Altogether the 90 cm deep soil profiles represent the main rooting zone for crops. Subsequent samples of the soil water and mineral nitrogen content were taken layer-wise and twice a year from 2000 to 2002.

Geo-referenced grain yields were obtained from a combine harvester yield monitor and processed within a 25 m radius around the sampling locations according to Jürschik (1999) providing the average yield around the grid and the standard deviation. While on the western side of the field (transects A to C) the nitrogen fertilization was uniform over the three seasons (at 30 grid points), the N doses applied on the eastern side of the field (transects D to F) were variable (also at 30 grid points). Therefore, the variable amounts of fertilizer are provided for each grid point according to the application map. Winter wheat, winter wheat and triticale were cultivated in 1999/2000, 2000/2001, and 2001/2002, respectively. The data set used in the present study is exclusively described in Wallor et al. (2018, in press; data set available online at doi: 10.4228/ZALF.2003.327).

2.1.2. *Foggia*

The field is located in Foggia (41° 27′ N, 15° 36′ E, 90 m above sea level) in southeast Italy. The results of a three-year research concerning the spatial and temporal variability of attributes related to yield and quality of durum wheat production were reported by Diacono et al. (2012). The trial on rainfed durum wheat was carried out on a 12-ha field cropped with *Triticum durum*, cultivar Gargano. The soil is a deep, silty-clay mesotrophic Vertisol of alluvial origin (IUSS Working Group WRB, 2015). The climate is characterised by hot and dry summers and precipitation mainly occurs in the winter months. One hundred georeferenced measurements of aboveground biomass (t ha^{-1}), yield (t ha^{-1}), protein and gluten contents of grain (% dry matter) were taken at harvest in three crop seasons (2005/2006, 2006/2007 and 2007/2008). In addition, LAI ($m^2\,m^{-2}$) was observed threefold during the first growing season for model calibration. For this study, we considered the hand-harvested variables monitored at the 100 georeferenced locations in each year collecting five aboveground samples of 1 m^2 of wheat plants. A conventional tillage in a continuous durum wheat system was applied: ploughing after harvest to 30 cm soil depth followed by secondary shallow tillage with a tine cultivator and harrow for seedbed preparation. Total nitrogen application of 90 kg N ha^{-1} was partitioned at the rate of 1/3 before sowing (incorporated by disk harrowing as ammonium phosphate) and 2/3 at the beginning of wheat tillering (as ammonium nitrate). Wheat harvest was carried out with a combined straw chopper. Hence, the chopped straw was spread on the soil surface during harvest. Initial soil properties were determined at 100 sampling points arranged in a regular grid. Soil sampling was conducted at the two upper soil layers of 0–20 cm and 20–40 cm. The air-dried samples were analysed for sand and clay contents by sedimentation analysis after chemical dispersion and organic matter oxidation (Deshpande and Telang, 1950). Organic matter content was determined by the method of Walkley and Black (1934). Results of the described soil analysis are shown in Fig. 3 as spatial distribution of soil organic matter and clay content. Bulk density was determined using the pedotransfer functions published in Hollis et al. (2012). For step a of model calibration (Section 2.3) the soil variables of the bottom layer of 40–200 cm were set equal to the averages of the two upper soil layers.

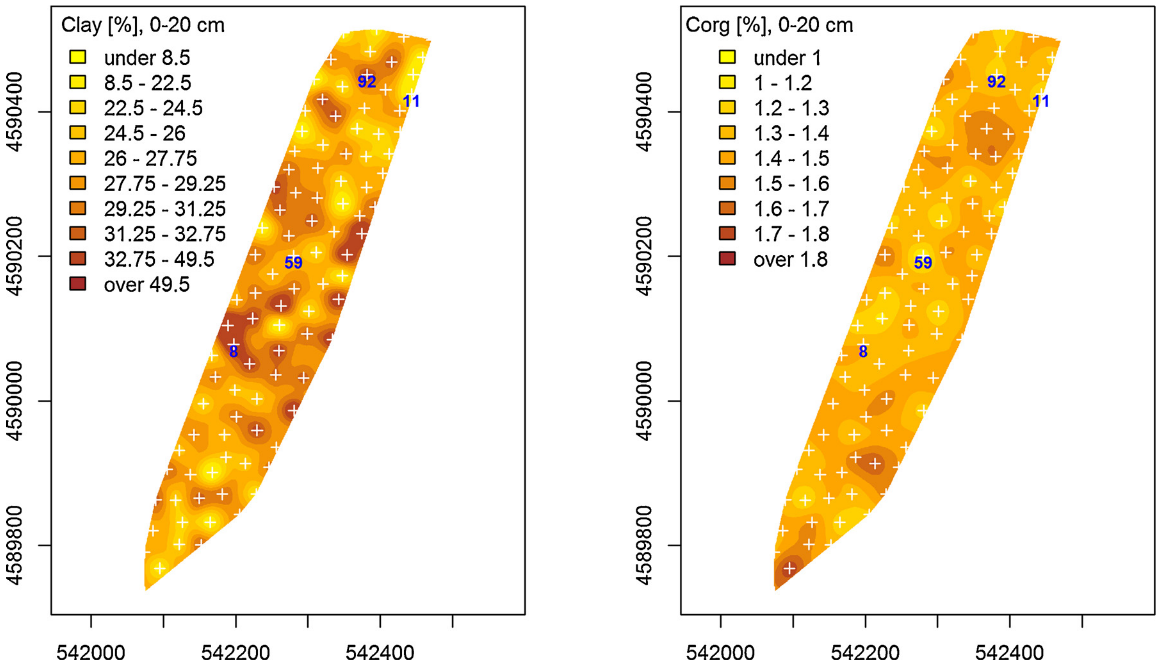

Fig. 3. Spatial distribution of clay content (0–20 cm) and organic carbon (C_{org}, 0–20 cm) at the field Foggia, Italy, and grid points of the sampled grid (n = 100).

2.2. Ensemble of crop models

Eleven agro-ecosystem models were tested with respect to their sensitivity to spatial variability. We classified them according to Asseng et al. (2013) as presented in Table 3. Main differences between the models exist according to the approaches implemented for simulating the crop development and the carbon and nitrogen dynamics in the soil. For example, the number of considered organic matter pools ranges from two (CO, HE) to even ten (LI). Yield formation is predominately represented by partitioning of assimilates along crop development during reproductive stages (CO, DA, DR, HE, LI, MO). A fixed harvest index is implemented in AS, CR, DS, DT and EP, but is modified by potential environmental constraints.

2.3. Model calibration

For both sites the first year of observations was used to calibrate the models. We considered three steps (a, b, and c) in a gradual calibration approach. In every step, additional data were provided successively to improve model calibration. By doing so, we were able to identify the site sensitivity of each agroecosystem-model at different levels of calibration intensity. At the baseline (step a) soil texture information was provided for every grid point at Beckum and Foggia since models use different procedures to derive their soil hydrological parameters from textural information. Due to the fact that most of the models consider soil profiles to a depth of 2 m soil information from the deepest investigated layer was adopted for the bottom layer (Beckum) or soil information of examined soil layers was averaged and passed to the bottom layer (Foggia). Additionally, we supplied soil organic matter of the ploughing layer (0 to 30 cm), initial values for soil mineral nitrogen content and the soil water content after harvest of the previous crop (winter oilseed rape) and layer-based estimated stone content for the Beckum site. For the field Foggia, soil organic matter and total nitrogen content was provided for the layers 0 to 20 cm and 20 to 40 cm. The initial soil water content was assigned to field capacity at the beginning of simulation (01.01.2005). In order to derive the initial mineral nitrogen content from the supplied total nitrogen content we suggested to consider a spin-up year and to start the simulation one year earlier (01.01.2004). The relevant agronomical and climatic input data were provided for both fields. To give the opportunity to calibrate model-specific plant parameters the maximal observed yield of the first (and third season for triticale at Beckum) season without providing the corresponding grid point was specified and available phenology data were transferred. In the second step (step b) we provided measured water field capacity and wilting point parameters for each layer and grid point to compare crop model behaviour and response based on the same hydrological properties excluding the influence of their model inherent parameter estimation. For the final step (step c) observed data from all grid points of the first year were distributed including yield, soil water and soil mineral nitrogen contents for the Beckum site, and yield, biomass and LAI for the Foggia site. The corresponding data of the second and third year were used for model validation. We expected to see a continuous improvement of models' performance from step to step.

Table 2
Range of clay, silt and sand content per soil texture class (Ad-hoc AG Boden, 2005, German Soil Taxonomy).

Soil texture class	Clay [%]	Silt [%]	Sand [%]
Sl2	5–8	10–25	67–85
Sl3	8–12	10–40	48–82
Sl4	12–17	10–40	43–78
Ls3	17–25	30–40	35–53
Ls4	17–25	15–30	45–68
Lt2	25–35	30–50	15–45
Lt3	35–45	30–50	5–35

2.4. Validation of model outputs

Model evaluation for the last two years was conducted using the programme R (Core Team R, 2014) and the package *hydroGOF* (Zambrano-Bigiarini, 2014). For the visualisation of our results we principally used the packages *ggplot2* (Wickham, 2009) and *openair* (Carslaw and Ropkins, 2012). Evaluation of model performance at the Beckum field focussed on the simulated results for yield, soil water content (WC) and soil mineral nitrogen content (N) of the soil profile (0–90 cm). At Foggia we were forced to concentrate on plant-related validation data (final grain yield and biomass, LAI) due to missing

Table 3
Main modelling approaches of eleven process models used in the MACSUR study on spatial variability.

Model (abbrevation)	Leaf area / light interception [a]	Yield formation [b]	Phenology [c]	Environmental constraints involved [d]	Water dynamics [e]	Evapotrans-piration [f]	Soil CN-model [g]	Climate input variables [h]
APSIM (AS)	RUE	HI(Gn)/B	T/DL/V	W/N	C	PT	CN/P(3)/B	R,Tx,Tn,Rd
CropSyst (CR)	RUE	HI/B	T/DL/V	W/N/H	C/R	PM	N/P(4)	R/Tx/Tn/Rd/RH/W
CoupModel (CO)	RUE	Prt/B	T	W/N/A	R	PM	CN/P(2)	R/Ta/Rd/RH/W
Daisy (DA)	P-R	Prt	T/DL/V	W/N	R	PM	CN/P(6)/B	R/Tx/Tn/Rd/e/RH/W
DSSAT/Cropsim (DR)	RUE	Prt	T/DL/V	W/N	C	PT	CN/P(4)/B	R/Tx/Tn/Rd
DSSAT 4.6 (DS)	RUE	HI(Gn)/B	T/DL/V	W/N	C	PT	CN/P(4)/B	R/Tx/Tn/Rd
DSSAT 4.6 (DT)	RUE	HI(Gn)/B	T/DL/V	W/N	C	PT	CN/P(4)/B	R/Tx/Tn/Rd
EPIC (EP)	RUE	HI	T/V	W/N/H	C	P/PM/PT/HAR	N/P(5)/B	R/Tx/Tn/Rd/RH/W
HERMES (HE)	P-R	Prt	T/DL/V/O	W/N/A	C	PM/PT	N/P(2)	R/Tx/Tn/Rd/e/RH/W
LINTUL-SIMPLACE (LI)	RUE	Prt/B	T/DL/V	W/N/H	D	PM	CN/P(10)/B	R/Tx/Tn/Rd/W
MONICA (MO)	P-R	Prt	T/DL/V/O	W/N/A/H	C	PM	CN/P(6)/B	R/Tx/Tn/Rd/RH/W

[a] RUE, radiation use efficiency approach; P–R, gross photosynthesis - respiration.
[b] HI, fixed harvest index; B, total (above-ground) biomass; Gn, number of grains; Prt, partitioning during reproductive stages.
[c] T, temperature; DL, photoperiod (day length); V, vernalization; O, other water/nutrient stress effects considered.
[d] W, water limitation; N, N limitation; A, aeration deficit stress; H, heat stress.
[e] C, capacity approach; R, Richards approach; D, Darcys approach.
[f] P, Penman; PM, Penman-Monteith; PT, Priestley –Taylor; TW, Turc-Wendling; MAK, Makkink; HAR, Hargreaves; SW, Shuttleworth and Wallace (resistive model); EB, energy balance.
[g] CN, CN model; N, N model; P(x), x number of organic matter pools; B, microbial biomass pool.
[h] Cl, cloudiness; R, precipitation; Tx, maximum daily temperature; Tn, minimum daily temperature; Ta, average daily temperature; Td, dew point temperature; Rd, radiation; e, vapor pressure; RH, relative humidity; W, wind speed.

Table 4
Overview of model results considered for evaluation.

Model	Beckum			Foggia		
	Yield	Soil nitrogen	Soil water	Yield	Biomass	LAI
AS	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
CR	a,b,c	–	a,b,c	a,b,c	a,b,c	a,b,c
CO	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
DA	a,c	a,c	a,c	a,c	a,c	a,c
DR	a,b,c	–	–	a,c	a,c	a,c
DS	a,b,c	a,b,c	a,b,c	a,c	a,c	a,c
DT	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
EP	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
HE	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
LI	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c
MO	a,b,c	a,b,c	a,b,c	–	–	–

information about soil state variables. We applied the following methods to compare our observed values with simulated values: the root mean square error (RMSE; Fox, 1981):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (S_i - O_i)^2}{n}} \tag{1}$$

and the mean error (ME):

$$ME = mean(S_i - O_i) \tag{2}$$

where n is the number of samples and S_i and O_i are the simulated and the observed values. The ME is most suitable to determine if the simulated output under- or rather overestimates the observations. Additionally, we calculated the absolute correlation coefficient according to Spearman (r) to evaluate the relation between simulated and observed values. In general, we evaluated the model outputs annually per each calibration step and based on each field and observed variable (RMSE, ME, r) as well as over the entire simulation period (r). To check the overall consistency of a model we summarised the correlation coefficient determined for each observed variable and divided the result by the number of observed variables. By doing so, we were able to verify the overall performance of each model irrespective of the different dimensions of observed variables. The overall model consistency was calculated solely for the Beckum site, where the observed variables include crop and soil information. On the whole, the three-year simulation period at each site takes into account carry-over effects and allows for consideration of the 1:2 relation between calibration and evaluation data. As suggested by Asseng et al. (2013) we included the model ensemble mean (MM) in our evaluation of model outputs in order to realise the overall tendency in stepwise calibration.

3. Results

During the modelling exercises described, individual outputs were excluded from evaluation for a variety of reasons; i) the simulated output did not correspond to the given reference values, ii) the calibration step was considered as ineffective by modellers, or iii) model output was simply not formatted correctly. However, the output of at least nine models could be evaluated per observed variable and field (Table 4).

Figs. 4 and 5 show examples of the effect of calibration steps at selected grid points for both sites for selected models. It becomes obvious, that the site conditions and, hence the yield performance varies between the grid points. Furthermore, the last calibration step does not necessarily lead to the best model fit at single grid points.

3.1. Models' performance at Beckum, Germany

3.1.1. Inter-annual dynamics of yield, soil water and soil nitrogen

The range of observed yields at Beckum reached from 5.78 t ha^{-1} to 7.97 t ha^{-1}, 5.46 t ha^{-1} to 7.96 t ha^{-1}, and 4.82 t ha^{-1} to 7.66 t ha^{-1} in the year 2000, 2001 and 2002, respectively. The lower yield potential in the year 2002 was caused by the cultivation of triticale after two years of winter wheat. The trend in the inter-annual variability of soil water contents was sufficiently well reflected by AS, CO, CR, DS, HE LI and MO. DA shows a similar trend but overestimated soil water

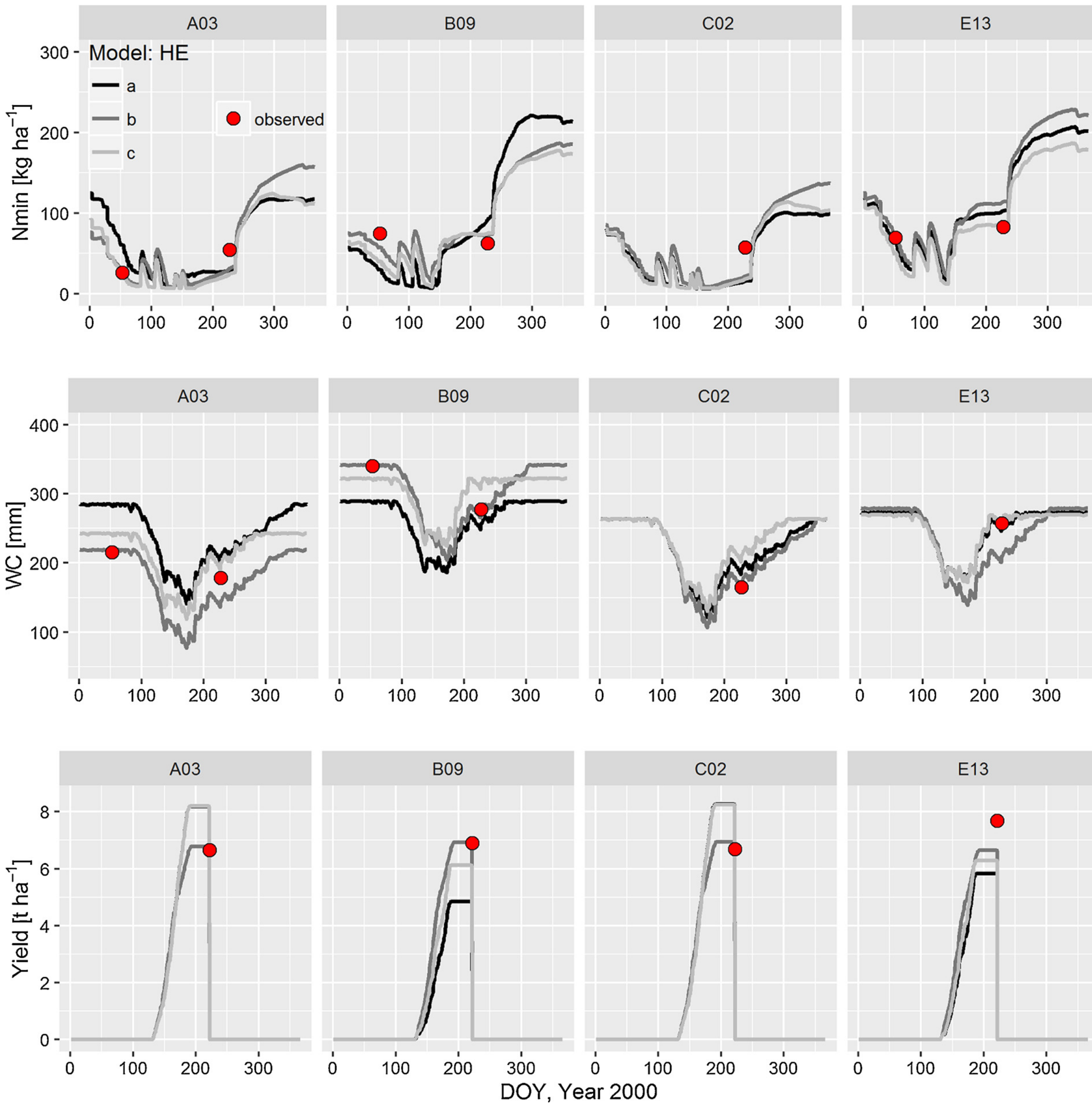

Fig. 4. The effect of calibration steps at selected grid points for Beckum and the HERMES model (grid point A03: Sl4; grid point B09: Sl2; grid point C02: Sl3; grid point E13: Lt2; averaged for the soil layer 0–90 cm).

contents in all years, while DT underestimated soil water. Trends of inter-annual yield variability was well reproduced by AS, CO, DT, and HE and the multi-model mean, while EP showed contrasting annual trends as consequence of the bad soil water simulations during the calibration year. However, differences were more pronounced than observed for AS, CO and DT. The within field variability of yields was only partially reproduced by the models DA, LI and MO. (Fig. 6).

The inter-annual differences in the soil nitrogen content were well depicted by the models AS, DS, HE, LI, and MO. CO and EP reproduced the range of observed soil nitrogen contents in the year used for calibration, but clearly overestimated the range in the following two years of the simulation period. Conversely, DA, DT, and MO did not match with the observed range in the year 2000. In the following years, DA substantially exceeded the observed range, while DT met the median of observed soil nitrogen.

3.1.2. Model validation based on year and calibration step

When we first consider the year of model calibration (2000) in relation with the observed soil variables (N, WC) a continuously increasing correlation coefficient (r) reflected the stepwise improvement of model output generated by AS, CO, HE, DA, and DS (Fig. 7, Supp. 1 and 2). In connection with a change of RMSE and ME the highest calibration effect was achieved for AS, DA (N), and DS (WC). For DS (N), CO, and HE both RMSE and ME were only slightly affected or even increased to a small extent. Despite a considerable improvement of r from step a to c regarding the WC simulation of DA, calculated errors increased during stepwise calibration. A clearly negative calibration effect was determined for DT (WC) reflected by increasing errors and a decreasing r. A small negative calibration effect also showed MO for N, while WC simulated by LI and MO as well as N simulated by EP remained unaffected by the calibration. Regarding the N simulation of LI, errors and r developed contrarily.

As shown in Fig. 7, the successful stepwise model calibration to meet the observed soil conditions did not automatically lead to an improved model performance regarding yield dynamics on the field.

The most consistent model fit was achieved by CO, DA, and HE

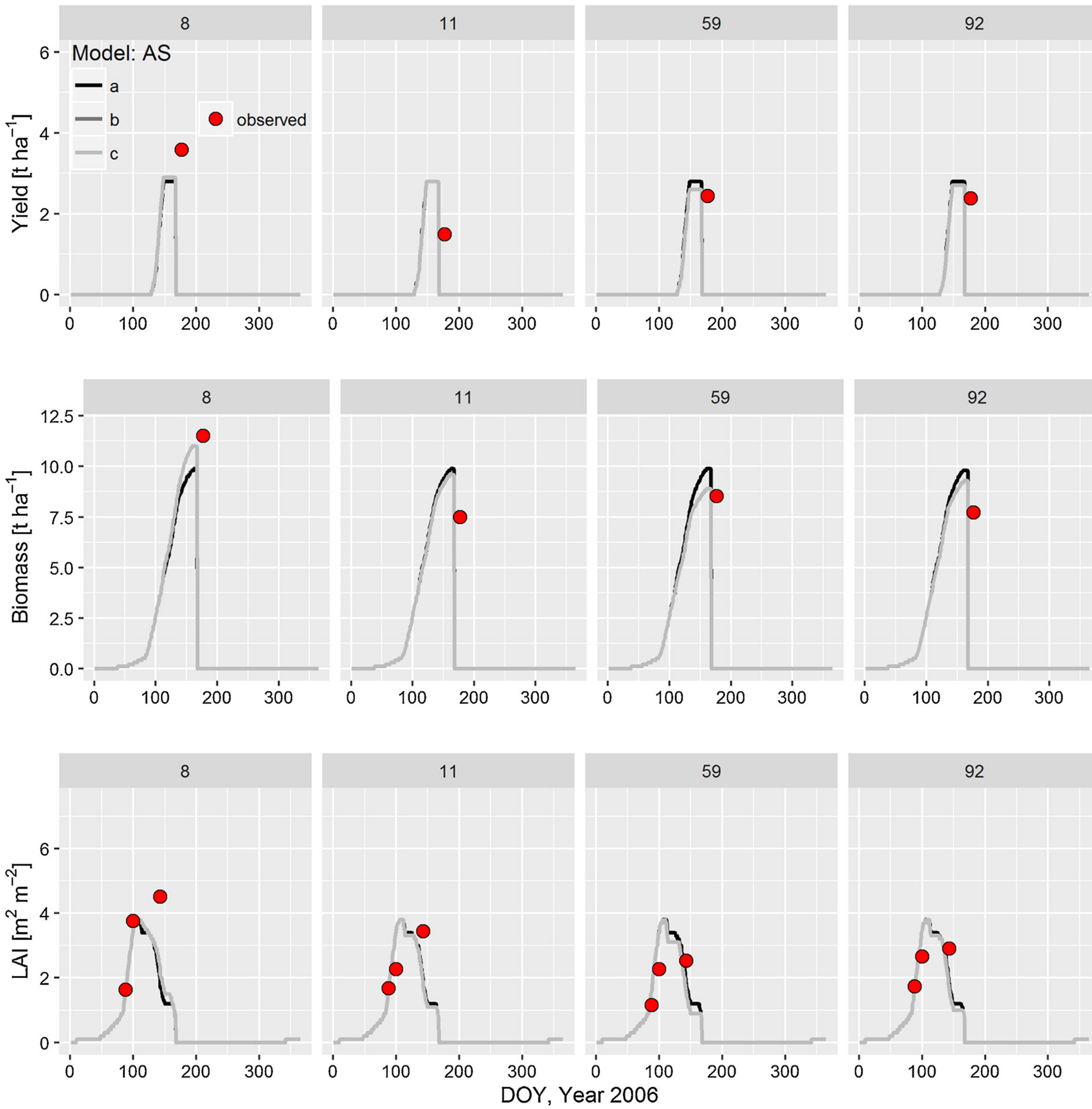

Fig. 5. The effect of calibration steps at selected grid points for Foggia and the APSIM model (grid point 8: 42% clay, 1.4% C_{org}; grid point 11: 12% clay, 1.2% C_{org}; grid point 59: 26% clay, 1.0% C_{org}; grid point 92: 35% clay, 1.2% C_{org} in the top soil layer).

resulting in a stepwise error reduction associated with a rise of r. No calibration effect on simulated yields could be determined for EP and LI, while yield simulation of DR, DS, DT, and MO deteriorated in the course of calibration intensity. Yields simulated by AS improved when regarding the calculated errors, but the reduced r suggested a declining relation between observation and simulation. In total, the models performed best in simulating the soil water dynamics. In the year used for model calibration, r ranged between 0.05 (DA, step a) and 0.89 (HE, step c) and RMSE between 27.02 mm (HE, step c) and 122.82 mm (DS, step a), respectively. The ensemble mean performed with r = 0.77 (step a) and r = 0.88 (step c). Also for the following simulation years soil water simulation resulted in a high r value for the MM of 0.68 (2001) and 0.72 (2002) and the final calibration step. Except for DT, model calibration produced consistent soil water dynamics over the years. The majority of the models simulated the soil nitrogen dynamics with an acceptable error reflected by a RMSE of 44.34 kg ha^{-1} for the MM in the year 2000. But relatively low values for r ranging between 0.10 (DS, step a; LI, step c) and 0.33 (HE, step c) suggested a limited relation between observed and simulated N. This pattern was not constant, and values of r increased during the following years used for model validation. A considerable improvement, consistent over time, was achieved for almost all models except EP, DS, and DT. The response of yield simulation on model calibration varied considerably between models and simulated years. For example, the successful model fit of CO and HE achieved after final calibration in the year 2000 declined again in the subsequent simulation period. Conversely, the clearly limited calibration effect obtained for the models AS, CR, DR, and DT resulted in a surprisingly good model fit for the yields observed in 2001 and 2002. No consistency in the calibration effect was observed for the model DA, neither when considering the steps nor the years of simulation. The models EP, LI, and MO performed relatively constant with low variations between steps and years.

3.1.3. *Overall model performance after final calibration*

The variable-based analysis of models' performance over the entire simulation period of three years is visualised in Figs. 8 and 9, and Supp.

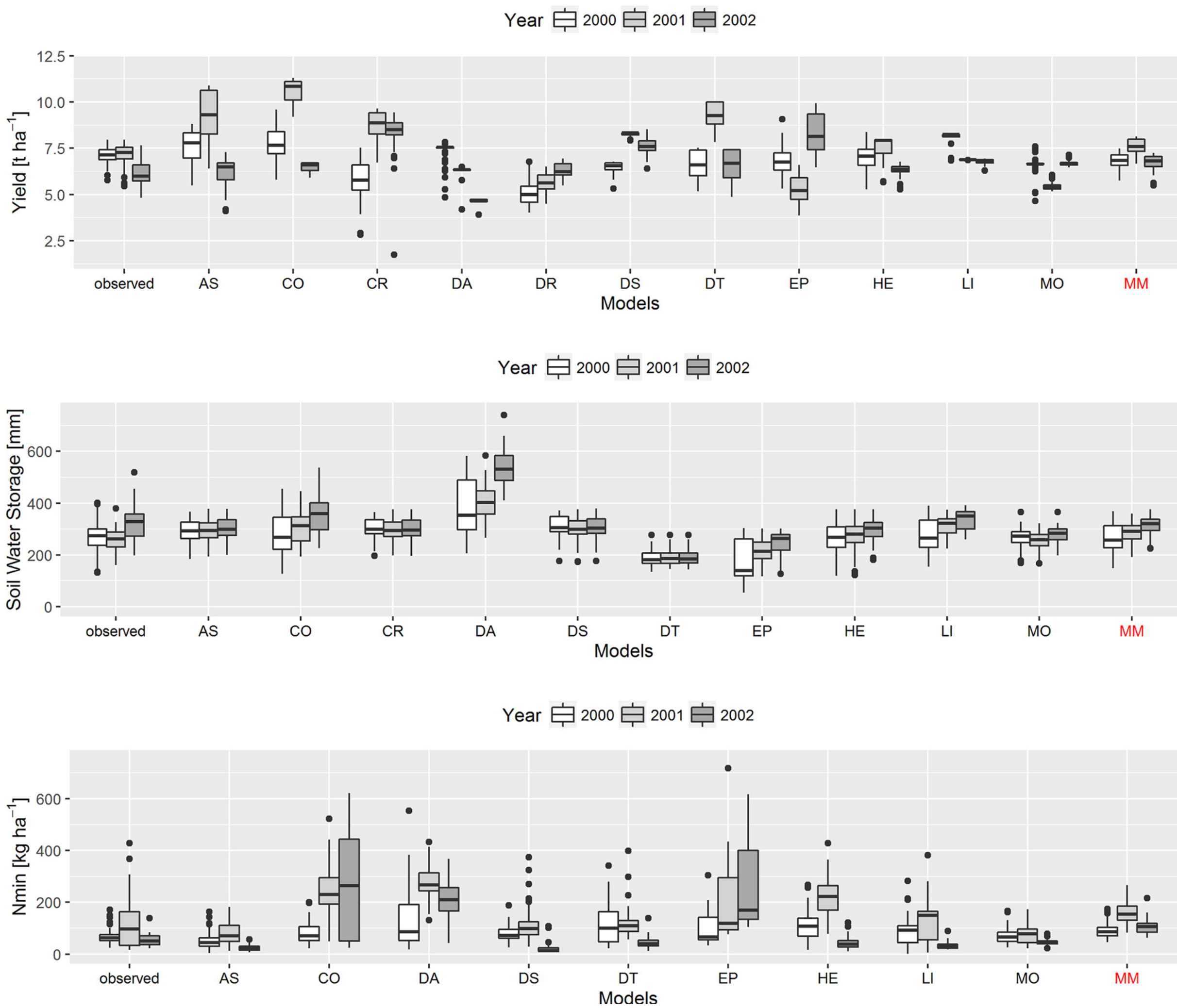

Fig. 6. Distribution of observed and simulated yields, soil water storage (0 to 90 cm soil depth) and soil nitrogen content (0 to 90 cm soil depth) per model and year for the field Beckum, step c; boxplots result from n = 60 grid points; MM = model ensemble mean.

3 considering only the final calibration step. With respect to the correlation plots of observed and simulated yields in association with the corresponding histograms some interesting model specifications were identified. Hence, the models DA, and LI produced different yield levels, reaching from two (LI) to three (DA). This resulted in extremely differing absolute correlation coefficients (r) ranging from 0.37 (LI) to 0.55 (DA). Negative correlations with varying intensity were found for the models CR (r = −0.28), DR (r = −0.46), DS (r = −0.08), EP (r = −0.55), and MO (r = −0.37), while the yields simulated with AS, CO, HE, and LI showed a relatively high positive correlation with the observed values. Consequently, the MM was only weekly correlated with the yield observations of all years and grid points (r = 0.20). Similarities between the models could be determined by calculating r for the several model outputs. For example, the yields simulated by AS, CO, DT, and HE correlated strongly with each other.

The yield output produced by the variants of the DSSAT model (DR, DS, DT) was similar solely between DS and DT (r = 0.70). Comparable yield simulation was also determined between CO and EP, DA and DR, DR, DS and LI, as well as between EP and MO. When considering the simulated soil water contents in step c) for the entire simulation period almost all models perform with an acceptable result of r.

An exception here was the output generated by DT and EP. The left-skewed distribution of WC simulated by DT resulted in a relatively low value for r. Except for DT, the simulated WCs of all models were correlated with each other (r > 0.30). A particularly strong relation was found with respect to the model results of AS, CR, and DS as well as of CO, HE, and LI. In addition, the soil water outputs of CR, HE, and MO were strongly correlated. As outlined before, the observed soil nitrogen dynamics were depicted by the models with a quite different accuracy (Supp. 3). The best model fit was achieved by HE, LI, and MO whose N output also correlated strongly with each other (r = 0.67 to 0.72). A very different distribution of simulated N from the observed values was found for CO, DA, DT, and EP resulting in values of r between 0.22 and 0.34. Similarities in the simulations of N were determined for AS, HE, and LI, DS and DT, DS and HE, as well as for DT and MO.

Finally, the overall model consistency calculated stepwise for the simulations conducted for the field Beckum are summarised in Table 5. As outlined before, the successful calibration of models with respect to the soil-related dynamics did not always lead to a sufficient simulation of yield development and vice versa (e.g. CR, DS, MO). Furthermore, the calibration effect differed between the models and the increasing amount of data for fitting the models did not necessarily improve the simulation (e.g. CR, DR, DT, EP, MO). As a result, the mean of the model ensemble did not represent the best model fit in the present exercise.

3.2. Models' performance at Foggia, Italy

3.2.1. Inter-annual dynamics of yield, crop biomass and LAI

For the field at Foggia solely crop-related variables were present for

Fig. 7. Stepwise calculated quality measures (r, ME, RMSE) of model performance per year regarding the measured yield [t ha^{-1}] at all grid points at Beckum.

model calibration and validation. The LAI was observed for the year of model calibration 2006 (Fig. 10). For the entire simulation period crop yield and biomass was measured at all 100 grid points. The observed range of LAI was between 0.67 $m^2 m^{-2}$ and 5.33 $m^2 m^{-2}$ and showed a normal distribution. The majority of observed LAI was successfully depicted by almost all models, except by CR, DT, EP, and LI. Thereby, the predominant number of models showed a skewed distribution of simulated values. CR, DT, and LI underestimated the observed leaf development, while EP considerable overestimated it. The ranges of observed yields and crop biomass were high as well as the inter-annual variability. In the year 2006 yields varied between 1.1 t ha^{-1} and 4.5 t ha^{-1} and biomass between 4.2 t ha^{-1} and 12.8 t ha^{-1}, respectively. One year later crop development declined and resulted in ranges of 0.2 t ha^{-1} to 2.7 t ha^{-1} for yield and 1.4 t ha^{-1} to 8.0 t ha^{-1} for biomass. In the last year of the simulated period the observed yields ranged between 0.4 t ha^{-1} and 4.5 t ha^{-1} and biomass ranged between 2.4 t ha^{-1} and 13.4 t ha^{-1}. According to Diacono et al. (2012) the average lowest yield of 2007 was likely due to differences in total precipitation during the growing season, with lowest values in the driest year of 2007. In fact, the 2007 seasonal rainfall (from September to May) was 372 vs. 482 and 431 for 2006 and 2008, respectively. Moreover another reason could be the highest mean maximum temperature (28.5 °C) recorded during the reproductive phase (April–July) of 2007.

Disregarding the distribution of observed crop variables, the inter-annual dynamic was successfully depicted by DA and HE, while most of the models produced an opposite dynamic with highest yields and biomass in 2007. A very low yield variability, annual and inter-annual, was simulated by AS, DR, DT, and EP. This also applied for the simulated biomass of EP. In total, DS considerably overestimated both crop variables, while the simulated biomass of DT and LI exceeded the measured values in the years 2007 and 2008. A less severe overestimation of biomass was generated by AS and CO. As a consequence, the multi-model mean did neither reflect the inter-annual dynamic nor the annual variability of crop development at Foggia.

3.2.2. Model validation based on year and calibration step

Corresponding to the results mentioned above and due to the fact that soil information was interpolated for a depth of > 40 cm based on the conditions observed in the upper layers, the model validation presents a very heterogeneous pattern. When we consider first the year used for model calibration (2006) a clear calibration effect was documented for the models CO, EP, HE, and LI associated with increasing values for r and a reduction of ME and RMSE, respectively (Supp. 4, Table 6). However, this effect was not necessarily stable during the rest

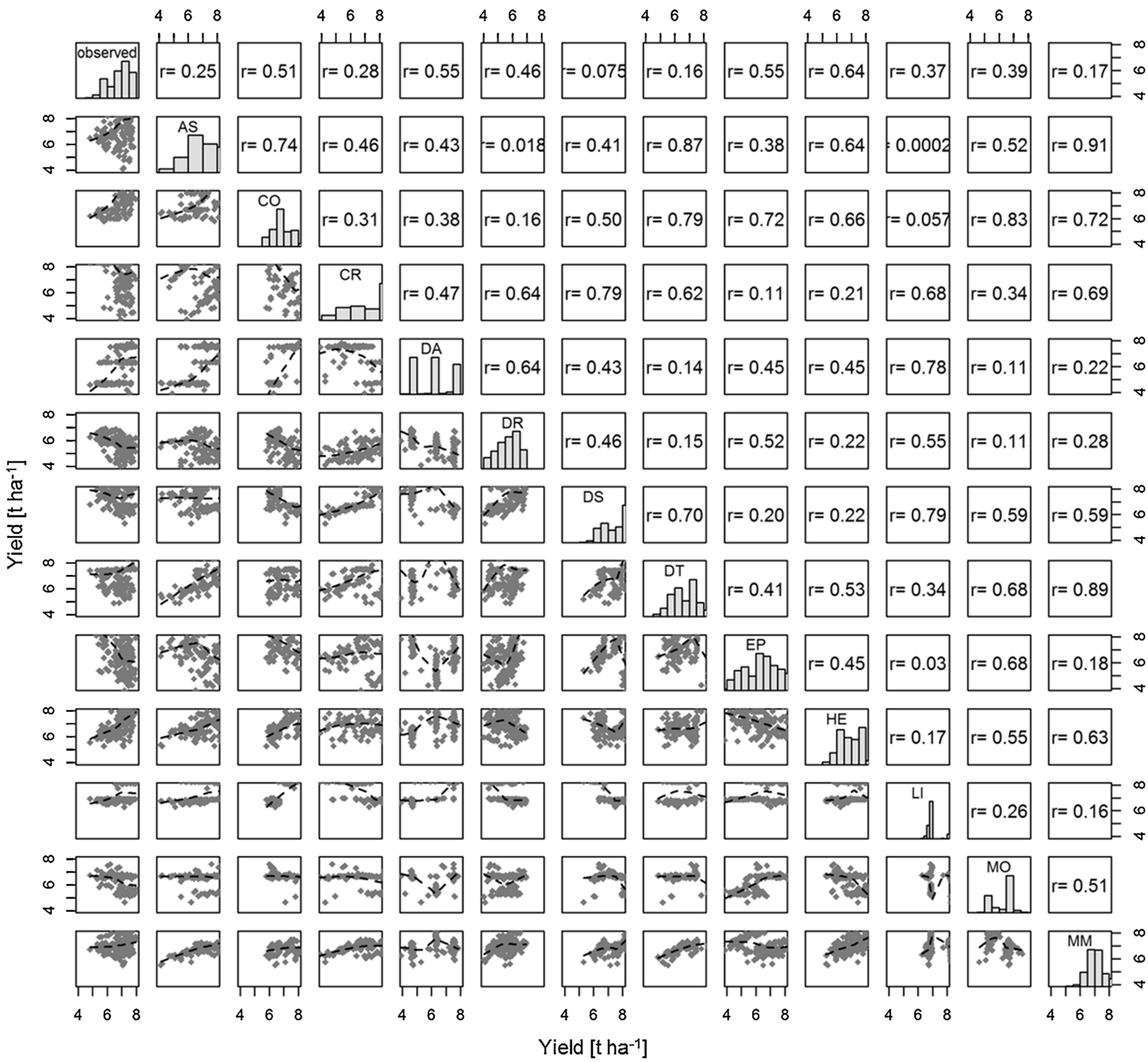

Fig. 8. Results of the final model calibration in step c according to the simulated yields for Beckum; r was calculated for all simulated years and grid points; left part: correlation plots, centre: histograms, right part: calculated r.

of the simulation period, except for EP. A reverse calibration effect on yield simulation was found for the models AS and CR, while more or less constant simulation results were produced by DA, DS, and DT. For the model DR the value for r kept constant during calibration, but ME and RMSE declined considerably.

Related to the weakly depicted inter-annual yield dynamic models' performance of the year 2007 was characterised by somewhat higher values for ME and RMSE associated with lower correlation coefficients. The lowest RMSE for this year was observed for DA and step c) (= 0.6 t ha^{-1}) accompanied with the lowest value for r (= 0.02). In the year 2008 the calculated values for r stayed entirely beneath a level of 0.15. The largest ME and RMSE were estimated for DS (step a and c), while CR, DA and HE tended to underestimate the observed yields with intensified calibration. A similar pattern according to the calibration effect was observed for the simulations of crop biomass. In the year 2006 highest values for r were achieved by CO, EO, HE, and LI during step c). Additionally, the RMSE was reduced during calibration regarding the output of DR, while CR, DA, and EP continuously underestimated the biomass production. This was also the case in the last year of the simulation period.

As for the simulated yields, RMSE increased for the majority of models in the year 2007 due to the observed inter-annual variability. The lowest error was determined for DA (step c) accompanied by relatively low values for r, while the highest deviation was found for LI (step a), DS (step a and c), and CO (step c). In total, the correlation coefficients estimated for the years 2007 and 2008 suggested an insufficient adjustment of all models to the biomass variability observed at the 100 grid points. This was generally reflected by r-values < 0.20. The results of the stepwise model validation per each year were extremely diverse. The interpretation of this diversity was complicated due to missing observations of soil variables. Models' performance according to the observed LAI was only weakly influenced by the calibration intensity (Table 6).

The majority of the models showed a successful initial performance which was reflected by r-values > 0.50 for all calibration steps. A negative calibration effect was observed for DR, DT, and EP, where r decreased continuously during calibration and RMSE increased, respectively. After the first step, CO, CR, EP, HE, and LI underestimated the observed leaf development. For the models CO and HE this underestimation changed to a slight overestimation in step c. The final calibration step led to negative MEs regarding the output of DA and DT. The most consistent simulation of LAI reflected by relatively constant values for r and RMSE was generated by AS and DS.

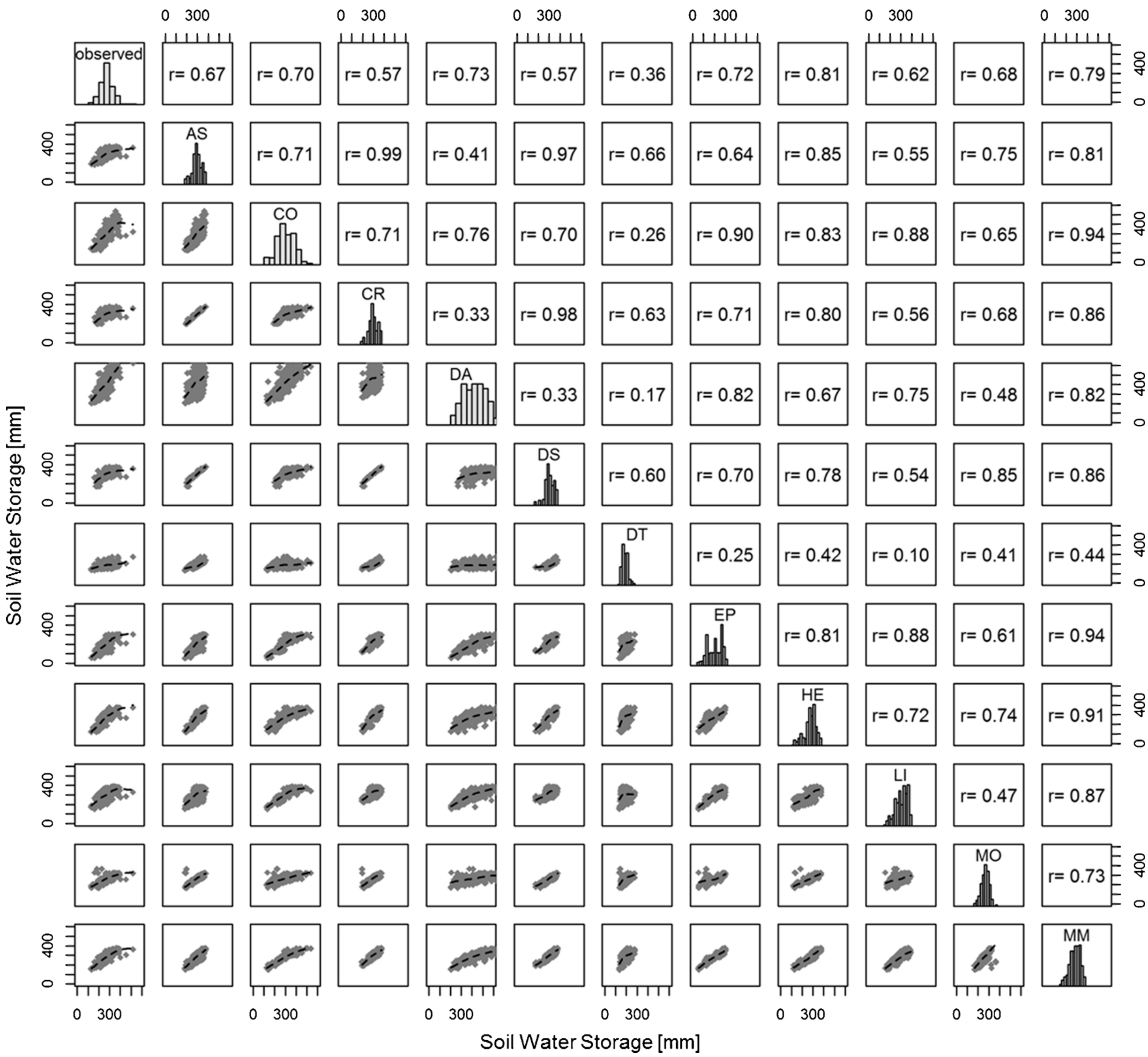

Fig. 9. Results of the final model calibration in step c according to the soil water simulations (WC) for Beckum; r was calculated for all simulated years and grid points; left part: correlation plots, centre: histograms, right part: calculated r.

Table 5
Overall model consistency, Beckum, step a, b, and c.

Model	Beckum		
	a	b	c
AS	0.32	0.25	0.47
CR[a]	0.23	0.10	0.14
CO	0.51	0.47	0.54
DA	0.24	NA	0.49
DR[b]	0.03	−0.20	−0.46
DS	0.24	0.33	0.33
DT	0.56	0.26	0.27
EP	−0.47	−0.47	−0.47
HE	0.26	0.32	0.65
LI	0.54	0.54	0.56
MO	0.42	0.42	0.36
MM	0.38	0.35	0.47

[a] CR simulated WC and yield, sum of correlation coefficients was divided by 2.
[b] DR simulated yield, correlation coefficient was not divided.

3.2.3. Overall model performance after final calibration

The results of models' validation presented up to here for the field at Foggia certainly determine the overall performance (Fig. 11, Supp. 4). Hence, a high negative correlation of simulated and observed crop variables was found for the models CO, CR, DT, and LI when considering the output of step c) for the whole simulation period. At least one negatively correlated crop output was simulated by DR (biomass) and DS (yield). A positive correlation, higher than 0.35, was determined for the simulated crop variables of AS, DA, and HE and the observed values, respectively. However, the crop variables simulated by EP showed a satisfactory correlation for biomass (r = 0.41) but a rather week model adjustment for yield (r = 0.15).

4. Discussion

4.1. Spatial variability and models' response

Soil properties play an important role in the soil-crop-atmosphere interactions since several bio-physical processes are influenced by them. Agro-ecosystem models, especially crop models, claim to consider the most relevant processes and inter-dependencies determining crop production, which suggests that they are suitable to assess effects

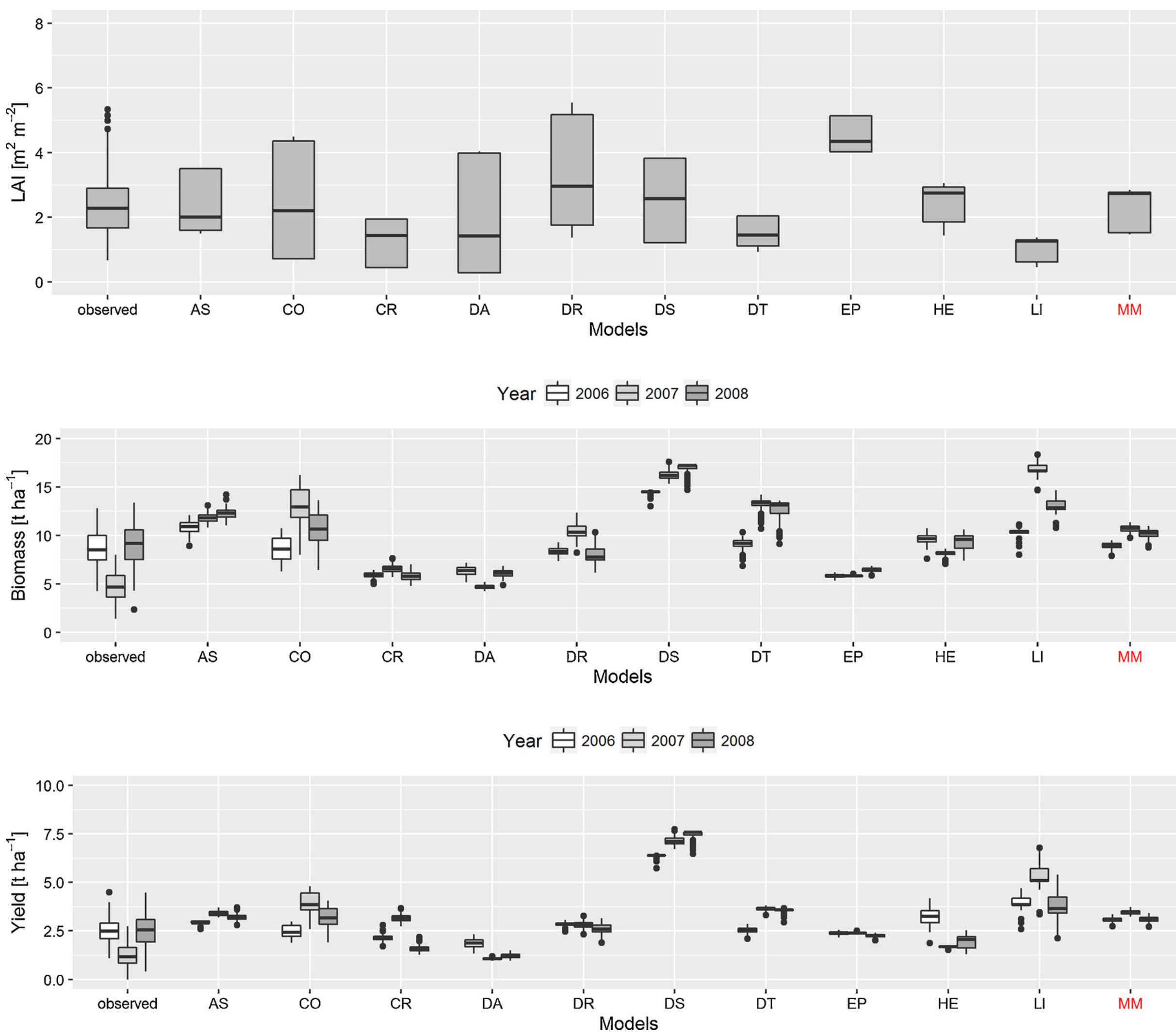

Fig. 10. Distribution of observed and simulated yields, biomass and LAI (2006) per model and year for the field Foggia, step c; boxplots result from n = 100 grid points; MM = model ensemble mean.

of management or changing climatic conditions at different scales. The two data sets used to investigate the capability of different models to reflect the effect of heterogeneous soil properties on spatial yield patterns showed differences regarding their availability of soil data and measured state variables. While the data set from Germany provided measured soil profile data down to 90 cm depth for each of the 60 grid points, the observed soil data set for the 100 grid points at Foggia was restricted to 40 cm depth. Whereas soil water contents and soil mineral N contents were observed twice a year at the investigated grid points of the German field in addition to annual yield patterns, the Italian data set provided only crop state variables (LAI, biomass, yield), which made it difficult for modellers to calibrate soil hydrological parameters during

Table 6
Stepwise calculated quality measures (r, ME, RMSE) of model performance regarding observed LAI [$m^2\,m^{-2}$] at all grid points, Foggia.

Variable	Model	2006								
		a			b			c		
		r	ME	RMSE	r	ME	RMSE	r	ME	RMSE
LAI	AS	0.53	−0.12	1.15	0.50	−0.06	1.08	0.50	−0.06	1.08
	CO	0.71	−1.01	1.48	0.71	−0.98	1.48	0.68	0.06	1.14
	CR	0.71	−1.37	1.52	0.71	−1.39	1.54	0.70	−1.12	1.30
	DA	0.69	0.10	1.08	NA	NA	NA	0.68	−0.48	1.29
	DR	0.53	1.05	1.61	NA	NA	NA	0.16	0.85	2.04
	DS	0.71	0.12	0.79	NA	NA	NA	0.71	0.12	0.79
	DT	0.68	1.70	2.93	0.54	−0.99	1.31	0.50	−0.88	1.24
	EP	0.71	−0.35	1.22	0.71	−0.35	1.22	0.18	2.10	2.35
	HE	0.68	−0.24	0.74	0.68	−0.29	0.75	0.64	0.09	0.70
	LI	0.68	−1.32	1.52	0.69	−1.34	1.53	0.69	−1.34	1.53
	MM	0.69	−0.15	0.75	0.71	−0.77	1.01	0.60	−0.07	0.70

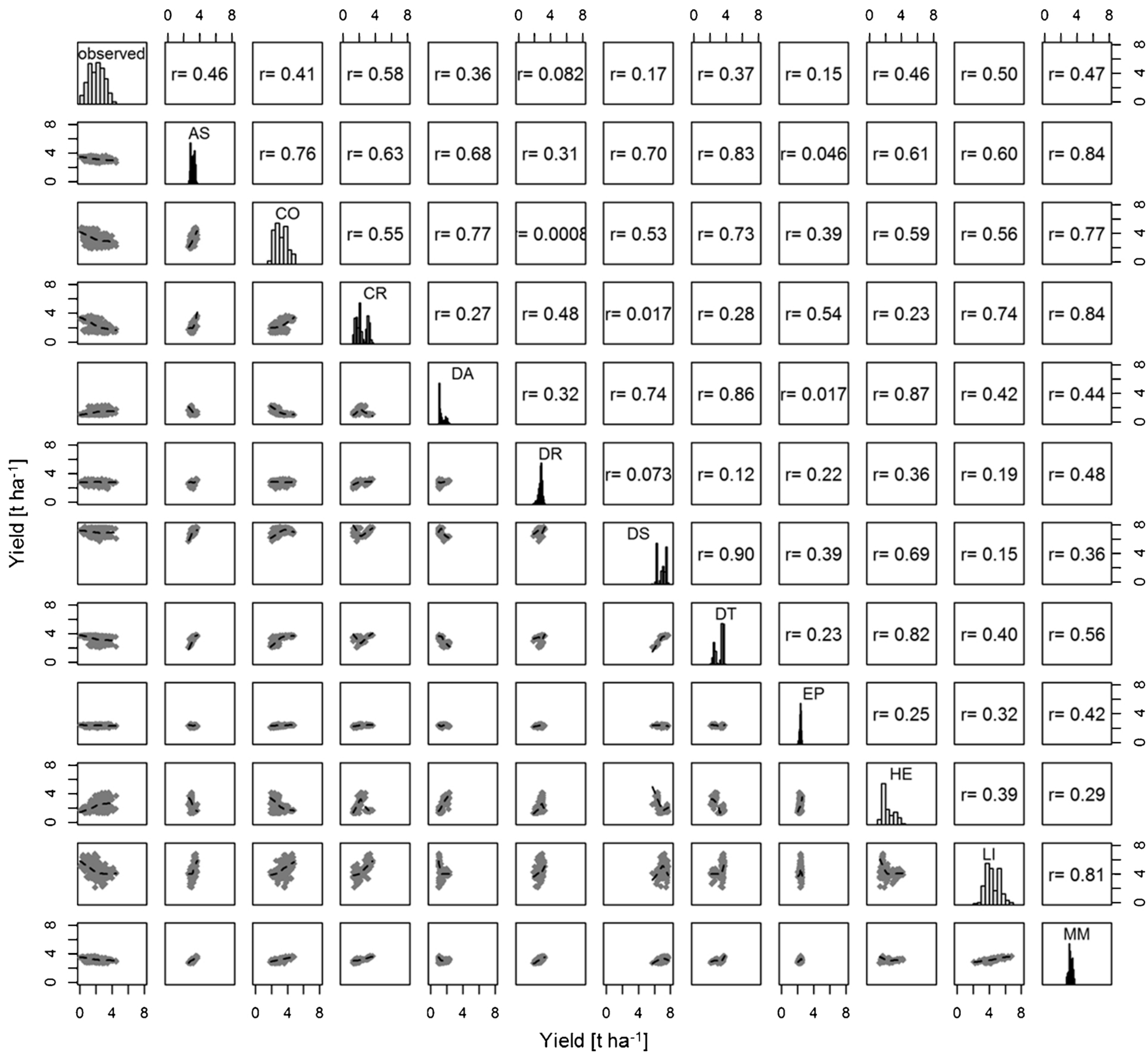

Fig. 11. Results of the final model calibration in step c according to the yield simulation for Foggia; r was calculated for all simulated years and grid points; left part: correlation plots, centre: histograms, right part: calculated r.

the calibration steps. While for the field at Foggia missing soil information below 40 cm was the highest source of uncertainty, for Beckum the sub-soil stone content in the southern part of the field, which was difficult to measure from auger sampling, created the main uncertainty in soil data. Addionally, not all models can account for stones in the profile. The small scale variability of the subsoil stone content is also reflected in the coefficient of variation (CV) of observed yield data within the buffer zone of 25 m radius around the grid points, which was obtained by the standard deviation of the single yield data from the combine harvester monitor within these zones. While the small scale variation within these zones was on average 9.6% regarding all three years, at some grid points, especially in the southern part and north of the forested field centre, CVs were greater than 13% (cf. Fig. 2). However, the majority of estimated CVs in the small scale were clearly below the experimental CV of ± 13.5% calculated by Taylor et al. (1999) across more than 300 field experiments (cf. Fig. 2).

Within field variability of gridded average grain yields was 14% for Beckum and 27% at Foggia/Italy, which corresponds well to findings of Joernsgaard and Halmoe (2003), who observed CVs for 82 Danish fields between 5 and 22% and a negative correlation between the yield level and the coefficient of variation. In order to clarify that yield variation is related to variable soil properties, especially caused by differences in soil texture and the related water holding capacity and plant available water, observed yields were assigned to the soil texture classes shown in Fig. 2. Such a grouping distinctly reduces the scattering of measured yields and, furthermore, clearly shows the impact of soil texture on yield performance of winter wheat, while this effect is less pronounced for triticale, which was cultivated in the growing season 2002 (Fig. 12).

An increase in winter wheat yields with rising clay contents up to the level of 17%–25% (Ls3, moderately sandy loam) becomes obvious, followed by a reverse decrease when the clay content exceeds 25% (Lt2, Lt3) or the silt content is less than 30% (Ls4). It is assumed that the present dynamic strongly depends on the relation between soil texture and soil hydrological properties as available water capacity first increases with higher clay contents and then drops off again when a certain amount of clay is reached (Ad-hoc AG Boden, 2005). In case of the Beckum field the drop off is intensified for the soil texture classes Lt2 and Lt3 due to the observed stoniness of the sub-soil and, hence the reduced rooting depth in the southern parts. This is also reflected in the distribution of measured soil water contents at a soil depth of 0–90 cm grouped by textural classes (data not shown). As already mentioned, the texture-based soil water availability as a driving factor for yield development plays a minor role for triticale, which generally is less responsive to water scarcity.

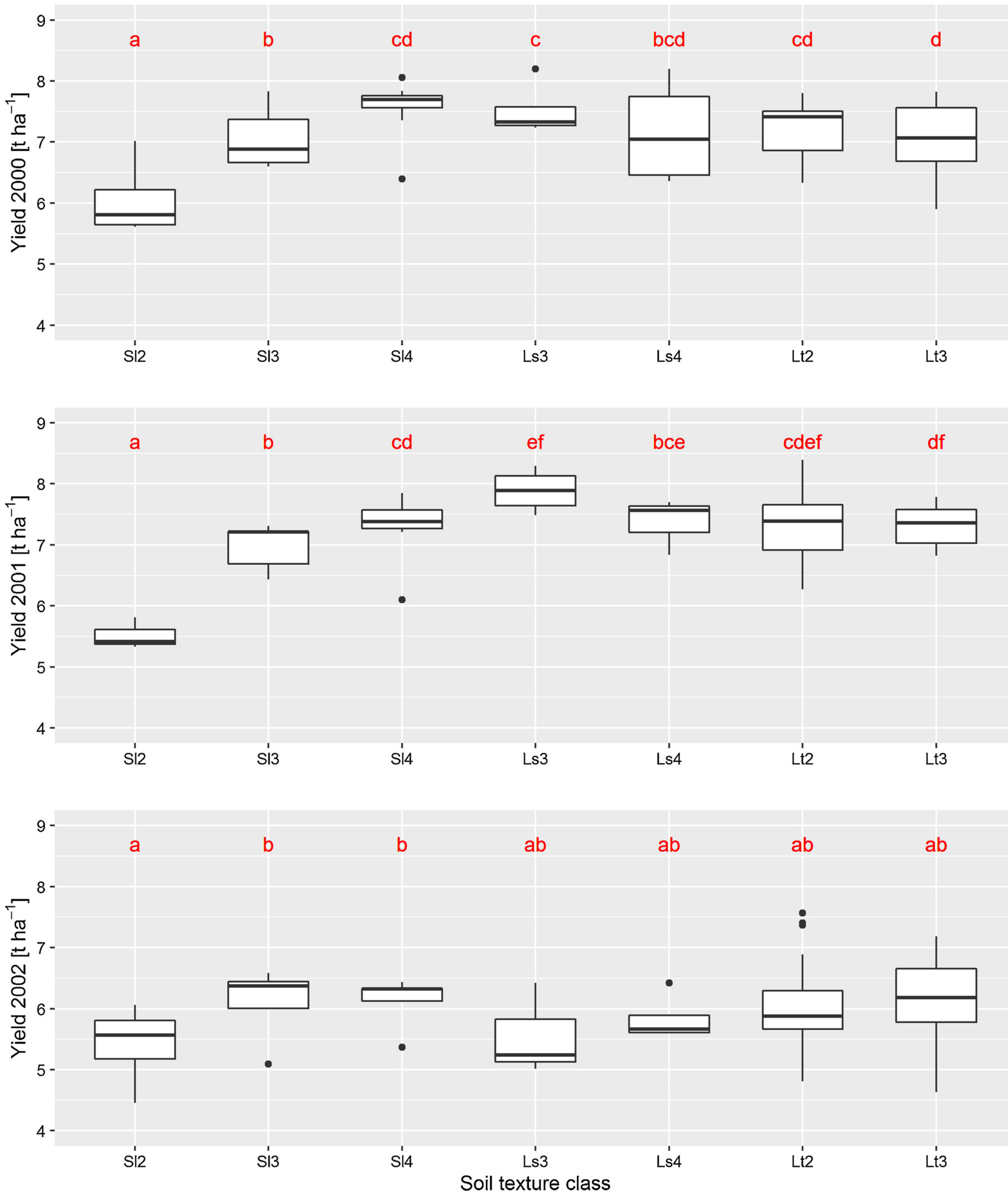

Fig. 12. Observed yields of winter wheat (year 2000 and 2001) and triticale (year 2002) grouped by soil texture classes (cf. Fig. 2, Table 2); values sharing the same letter are not significantly different (Mann-Whitney U-test; Mann and Whitney, 1947).

Uncertainty of crop models expressed as variance of results from large model ensembles was documented by several model inter-comparisons, e.g., within the AgMIP and MACSUR community (Asseng et al., 2013; Bassu et al., 2014; Martre et al., 2015). Results at single sites showed significant variability of models and results over various environments indicated, that no model turned out as a best performer across all sites or environments since models respond differently to environmental inputs. This can be caused by using different process response functions, e.g., for temperature on crop growth (Wang et al., 2017), evapotranspiration formula (Cammarano et al., 2016) or root distribution functions (Wu and Kersebaum, 2008). Additionally, even similar models derived from the same underlying model, but differing in parameterization can give quite different results (Folberth et al., 2016). Models' capability to capture the texture-related observations of winter wheat yields at Beckum differs extremely (Fig. 13). While AS consistently overestimates the texture-based yields, DR for example, strictly underpredicted them. A similar performance can be observed for DA, EP and MO with increasing clay contents in the soil, while CO shows a reverse response, although models' input has been adjusted to stoniness and rooting depth. A relatively weak texture response, accompanied by a broad distribution of yield values, presents CR. This is also true for DS and DT, considering the soil texture classes with more

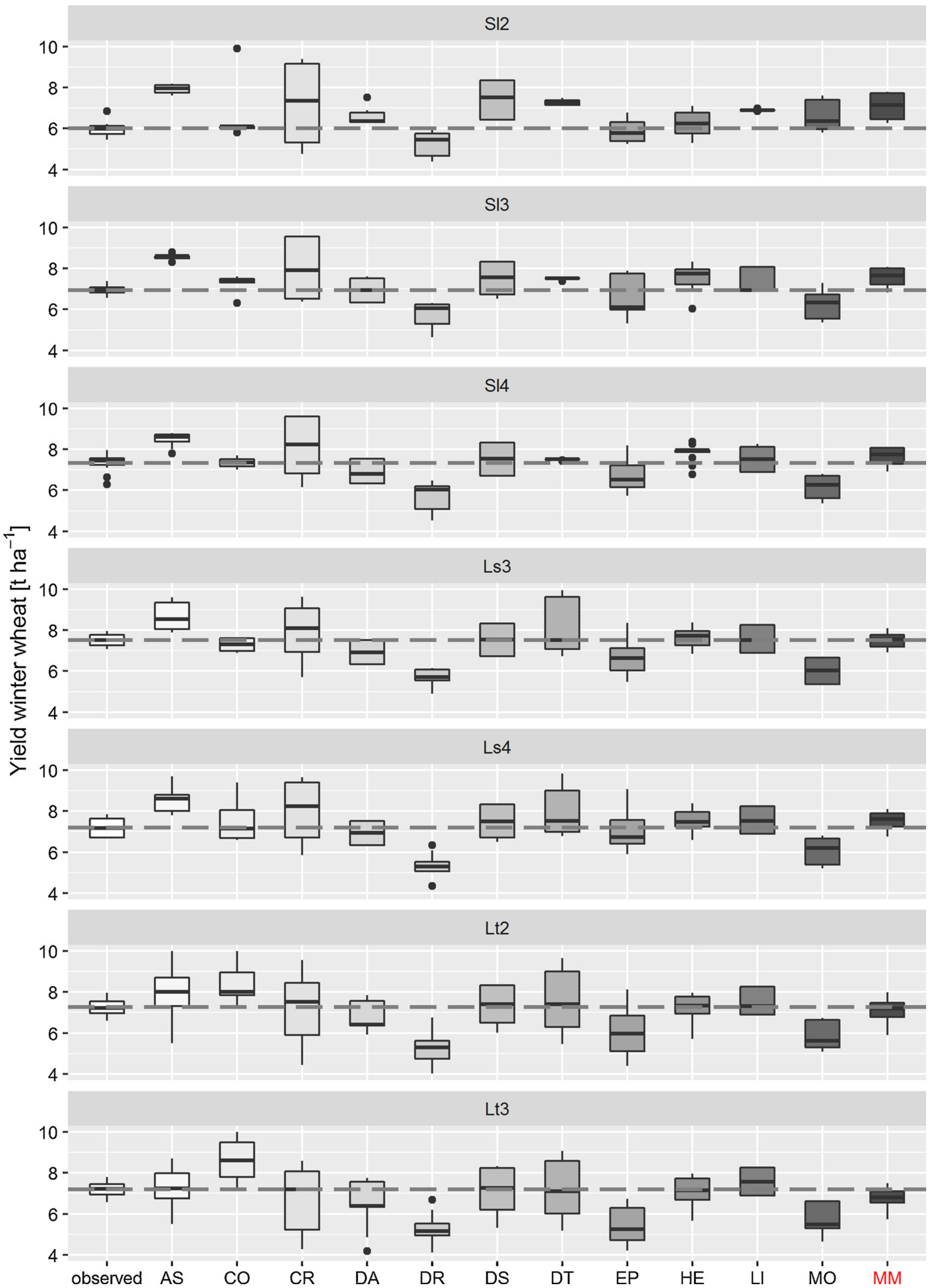

Fig. 13. Comparison of observed and simulated winter wheat yields (step c) at the Beckum site grouped by soil texture class; dashed line = average observed yield per texture class.

than 17% clay content. However, the yield level of the strongly sandy texture classes is overestimated by these models as well as by AS and to some extent by HE and LI.

It is supposed and outlined more detailed below, that these texture-based differences in models' yield simulation are – among other factors - strongly related to the individual approach used by models and modellers in order to parameterize soil hydrological conditions. For example, DA and LI both use the van Genuchten model to generate their site-specific water retentions curves but use different equations to simulate water dynamics (cf. Table 3). On the other hand, DT's and HE's water dynamics base on a more simple capacity approach, but soil hydrological characteristics are estimated differently. At this point it can be hardly distinguished how the individual soil hydrological estimations can be improved and what are the shortcomings. That needs to be looked at in more detail. In this context, it might be interesting that soil hydrological parameters have not been adjusted for the simulations with AS and LI as it was suggested (cf. Section 2.3). Additionally, the rooting depth in CR and LI was fixed to 1.0 m for all grid points regardless of the calibration step.

4.2. Models' consistency

The available water capacity is one of the most sensitive yield-affecting soil properties under rain-fed conditions with significant impact on crop model output, especially when water is a limiting factor (Aggarwal, 1995; Pachepsky and Acock, 1998; Wassenaar et al., 1999; Wong and Asseng, 2006). Simulation of soil water contents at Beckum were mostly improved for the models by using the standard

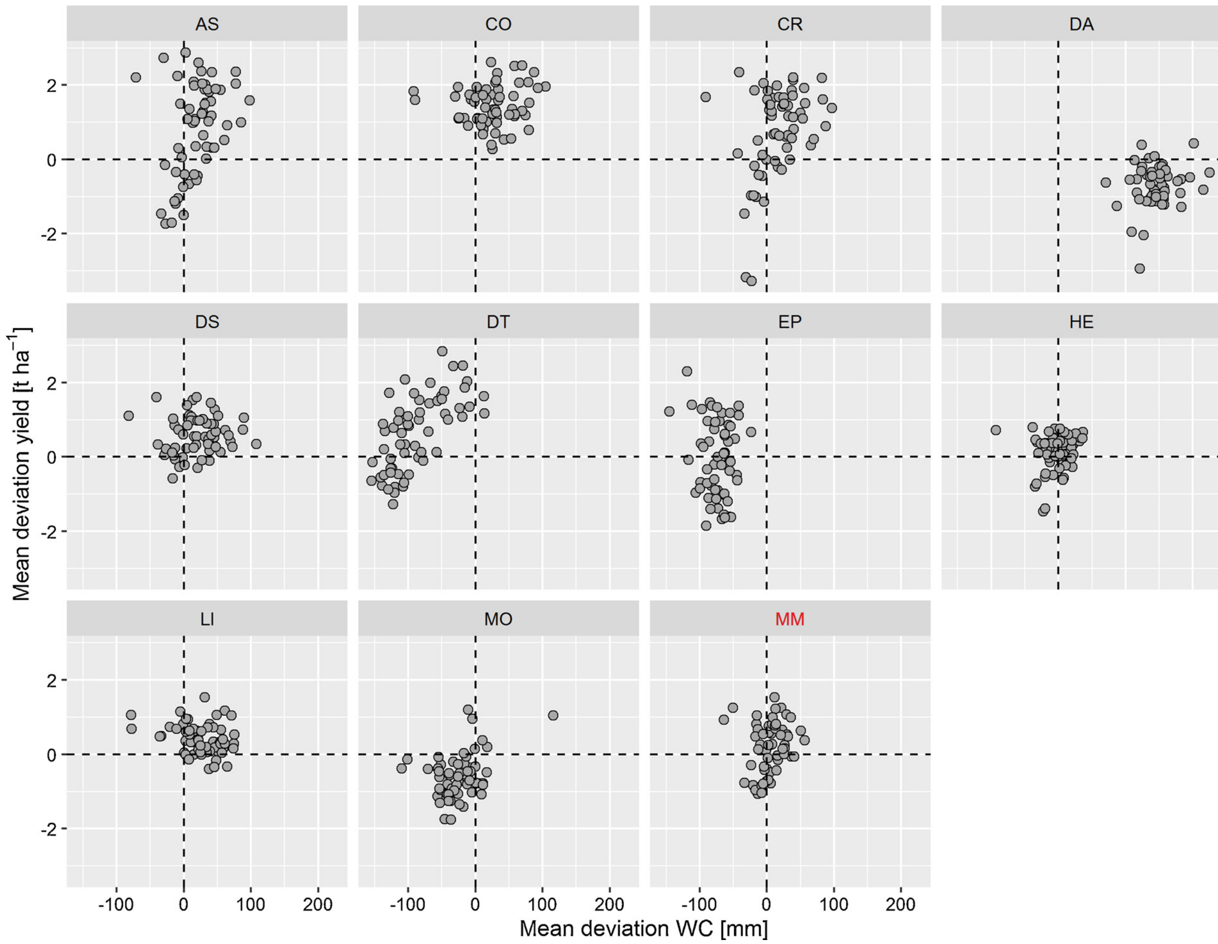

Fig. 14. Model-based relation between the mean deviation of simulated and observed yields and the mean deviation of simulated and observed WCs; mean deviation was calculated based on all observations per grid point of the entire simulation period, only step c was considered.

hydrological parameters derived from soil texture according to the German soil taxonomy (Ad-hoc AG Boden, 2005) instead of using the internally derived parameters of the models. Further fine calibration during step c for most models was mainly related to achieve a better agreement with soil mineral nitrogen content, but still improved soil water simulations (e.g. Supp. 1), which showed acceptable agreement for most models except DT after the final calibration step c. Although simulated water contents showed a good correlation to observed values for DA and EP, DA systematically overestimated water contents, while EP underpredicted them. The differences in soil water holding capacity suggest that soil water availability contributes mostly to the spatial patterns of crop production, whereas according to the relatively high level of soil mineral N in the root zone and sufficient amounts of N fertilizer, nitrogen limitation during the three growing seasons were not expected to limit crop growth significantly. Rötter et al. (2012) showed that MBEs of soil water content and grain yield of barley of a model ensemble were positively correlated under water limited conditions. Therefore, a positive relation between the deviations of simulated and measured water contents and yield deviations could be expected. However, not all models showed this in their simulations (Fig. 14). DT, for example, tended to underestimate the observed soil water contents and contrastingly overestimated the yield for the majority of grid points. Deviations of water content were strongly related to yield deviations (R^2 = 0.44) indicating that parametrisation of soil hydraulic properties was a significant source of error for the crop yield prediction and crop growth parameters have been adjusted on a wrong water supply. A similar inconsistency was for DA, which performed vice versa: lower yields were simulated associated with too high soil water contents. Since there was no correlation between the deviations of water and yield (R^2 = 0.04), the problem here seems to be a systematic overestimation of soil water content, e.g. by under-predicting evapotranspiration, and response function to water stress. A similar behaviour was observed in the study of Palosuo et al. (2011), where DA showed the best performance for wheat yield prediction on a sandy soil with overestimation of soil water.

AS and CR showed high deviations in yield predictions although water deviations were only slightly biased. Yield deviations showed a medium relation to soil water errors (R^2 of 0.14 and 0.15, respectively), which indicate that the response to soil water is over-predicted. HE and LI produced the most consistent results showing a relative small band of deviations for water and yields. The slight over-prediction of yields by LI was consistent to its slight over-prediction of soil water. CO's simulation was similarly consistent in this sense, but showed larger deviations for yield predictions. The same applies for MO, which constantly underestimated both the yields and the WC. This is in line with the cross-variable consistency listed in Table 5, which includes soil mineral nitrogen additionally. For all three models no correlations between the two variable deviations exist. These findings correspond to findings of a regional analysis on the effect of spatial data stratification strategies for the Federal State of North-Rhine-Westphalia/Germany (Zhao et al., 2016), who outlined that winter wheat yields simulated by AP and CR were highly sensitive to water holding capacity of soils, while CO, EP and MO showed only small and HE and LI moderate sensitivity. Although the site shows a high variability in soil properties the limiting effect of soil water is not pronounced due to relative high precipitation. Regarding the relation between small scale variability around the grid

points and the overall within field variability very high correlations cannot be expected. However, the differences are sufficiently big to expect a positive correlation between variables when models are consistent.

Surprisingly, some models (EP, CR, DR, DS, MO) showed even negative correlations between simulated and observed yields (Fig. 8). Although the experience of the model users has an influence on the results (Confalonieri et al., 2016), the results of two of three DSSAT (DR, DS) outcomes both showed the same strange behaviour, which can also be seen in Fig. 8 through their interrelation plot. For DA yield performance appeared at two to three distinct levels which might be an effect of soil property clustering, e.g. by texture classes. A similar behaviour can be observed for simulated crop yields of EP and LI, whereas MO showed nearly no response to the variable soil inputs (Fig. 7). Water is expected to be also the main factor at Foggia determining the productive performances of winter durum wheat (Ventrella et al., 2016) and the spatial patterns of crop yields and biomass. Since there were no data of measured soil water contents, soil hydrological parameters could not be properly calibrated for the grid points. Under the given dry conditions subsoil information on soil texture and plant available water would be essential to achieve a better performance of models. Baier and Robertson (1968) emphasized the strong impact of subsoil water availability at and after heading for wheat yield formation. McDonald et al. (2013) found that subsoil properties restricting root growth can limit water use and yield in many Australian soils. Kirkegaard et al. (2007) estimated a beneficial effect of available subsoil water even under moderate water stress at an efficiency of 59 kg ha^{-1} grain yield per mm subsoil water uptake and came to the conclusion that relatively small amounts of subsoil water can be highly valuable to grain yield. Missing information of subsoil properties might be the reason for the relative weak performance of all models for Foggia. Even an adjustment of subsoil parameters using the yield data of 2006 by HE did not improve the model performance in the validation years. Such an optimisation would require a number of spatial yield information from contrasting climatic seasons to find a unique solution, which fits to multiple years. The relative weak results for Foggia could be due also to a poor simulation capacity of the soil water stress effects on durum wheat growth and/or crop productivity. Furthermore, it is possible that models do not take into account the effects of extreme temperatures in spring months (Ventrella and Garofalo, 2016). In any case the results presented here underline the requirements for consistent data sets providing multiple soil and crop state variables in parallel and for several seasons for a proper model calibration (Kersebaum et al., 2015).

5. Conclusions

The presented results of the model inter-comparison investigating models' response to field-scale variability of soil properties on soil and crop state variables outline the importance of model calibration and further adaptation on the base of high quality data sets. In this context, high quality data sets should include multiple soil and crop state variables in parallel for several dates within one growing season (Kersebaum et al., 2015). Both data sets used in this study illustrate the manifestation of within-field heterogeneity in soil characteristics and its impact on local crop development. On the other hand, the way the data was collected and/or aggregated has a significant impact on data quality and, hence on model validation results. However, the annual and soil texture-based within-field variability of crop yield was only reflected by a few models. In relation with a highly sufficient depiction of heterogeneous soil water contents, that are supposed to mainly influence yield performance at both fields, this is a very remarkable result as it outlines the lack of linkage between soil and crop processes, which is expressed by the inconsistency of model outputs for some models. In this context, it is also interesting to note that increasing the scope of data in the step-by-step calibration process did not necessarily result in a better model outcome for single models and state variables. Bearing in mind that models showed a high uncertainty even at single sites in several model intercomparisons, it is not surprising, that the disagreement between models is excacerbated by increasing variability of environmental conditions caused by the heterogeneity within the fields. To some extent, this is certainly due to the very specific site conditions at some parts of the German field, notably the stone content, which affects the soil water processes in a way that models normally do not account for. The simulation results for the Italian field in particular showed that the majority of the models are only conditionally calibrated to extreme climates. Further adaptation is needed regarding heat and water stress. The overall rather weak models' performance for the Italian field is in particular also due to the interpolated and thus assumed subsoil information below 0.4 m soil depth. The often summarised view that the mean of all models gives the best result is not confirmed by the present study. While this applies to individual years and state variables, it cannot be generalised here. Regarding consistency across variables the model ensemble mean ranked at number five. Wallach et al. (2018) stated, that the statistical basis for the superiority of the ensemble mean is, that the model and environment-model interaction effects cancel out between models. However, if the error averaged over models shows a bias, this causes a bias in the ensemble mean as well. If one or few individual models have only a small bias, they outperform the ensemble mean. Therefore, the existence of bias tends to make the ensemble mean a worse predictor than the best model (Wallach et al., 2018). Although, it was not at the top as in most other studies looking at individual output variables it was still in the better half. The reason was the poor performance of some models showing even negative relations to the observed variables and a high bias. However, calculating the multi-model mean significantly reduces the deviations from the observed values produced by individual models. Therefore, it is still essential to select and use multiple models when making statements for larger scales as the quantification of the error caused by the aggregation of information presents a new challenge in view of the described results.

Acknowledgements

The authors are particularly grateful to Dr. Annamaria Castrignanò (CREA-AA) for providing the experimental dataset collected in Foggia and to Dr. Antonio Troccoli (CREA-CI) for the agronomic advice needed for the use of the dataset collected within the three-year project "Integrated System for Development of Southern Cereal Farming" (Sistema Integrato per lo sviluppo della Cerealicoltura Meridionale) fundedby Italian Ministry of Agricultural, Food and Forestry Policies (Program for southern Italy development: Research and Technological Innovation, Resolution CIPE 17/2003-1.1 and 83/2003). The presented model inter-comparison was established within the framework of FACCE JPI MACSUR2 and the contributing authors thank the National Ministries and the European Union for funding the projects (grant 0339740 and 031B0039C). EW and KCK were additionally funded by the I4S project within the BMBF BonaRes program (031A564I). MT and EP contribution has been funded through National Agency for Agricultural Research project no. QJ1610072 and by the Ministry of Education, Youth and Sports of the Czech Republic within the National Sustainability Program I (NPUI), grant number LO1415.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fcr.2018.08.021.

References

Ad-hoc AG Boden, 2005. Bodenkundliche Kartieranleitung, 5th edition. German Soil Taxonomy, Schweitzerbart, Stuttgart 438 pp.

Aggarwal, P.K., 1995. Uncertainties in crop, soil and weather inputs used in growth

models: implications for simulated outputs and their applications. Agric. Syst. 48 (3), 361–384.
Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3 (9), 827–832.
Baier, W., Robertson, G.W., 1968. The performance of soil moisture estimates as compared with the direct use of climatological data for estimating crop yields. Agric. Meteorol. 5 (1), 17–31.
Bassu, S., Brisson, N., Durand, J.-L., Boote, K.J., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Kim, S.-H., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Chang. Biol. 20 (7), 2301–2320. https://doi.org/10.1111/gcb.12520.
Cammarano, D., Rötter, R.P., Asseng, S., Ewert, F., Wallach, D., Martre, P., Hatfield, J.L., Jones, J.W., Rosenzweig, C., Ruane, A.C., Boote, K.J., Thorburn, P.J., Kersebaum, K.C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Heng, L., Hooker, J.E., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2016. Water use of wheat: simulated patterns and sensitivity to temperature and CO2. Field Crops Res. 198, 80–92.
Carslaw, D.C., Ropkins, K., 2012. Openair - an R package for air quality data analysis. Environ. Model. Softw. 27-28, 52–61.
Confalonieri, R., Bregaglio, S., Acutis, M., 2016. Quantifying uncertainty in crop model predictions due to the uncertainty in the observations used for calibration. Ecol. Modell. 328, 72–77.
Core Team R, 2014. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Austria, Vienna Accessed 03 July 2017. http://www.R-project.org/.
De Wit, C.T., Van Keulen, H., 1987. Modelling production of field crops and its requirements. Geoderma 40 (3–4), 253–265.
Deshpande, V.V., Telang, M.S., 1950. Pipet method of sedimentation analysis. Rapid determination of distribution of particle size. Anal. Chem. 22 (6), 840–841.
Diacono, M., Castrignanò, A., Troccoli, A., De Benedetto, D., Basso, B., Rubino, P., 2012. Spatial and temporal variability of wheat grain yield and quality in a Mediterranean environment: a multivariate geostatistical approach. Field Crops Res. 131, 49–62.
Easterling, D.R., Wallis, T.W., Lawrimore, J.H., Heim, R.R., 2007. Effects of temperature and precipitation trends on US drought. Geophys. Res. Lett. (20), 34.
Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartošová, L., Asseng, S., 2015. Crop modelling for integrated assessment of risk to food production from climate change. Environ. Model. Softw. 72, 287–303.
Folberth, C., Elliott, J., Müller, C., Balkovic, J., Chryssanthacopoulos, J., Izaurralde, R.C., Jones, C.D., Khabarov, N., Liu, W., Reddy, A., Schmid, E., Skalský, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P.J., Olin, S., Pugh, T.A.M., Ruane, A.C., Wang, X., 2016. Uncertainties in global crop model frameworks: effects of cultivar distribution, crop management and soil handling on crop yield estimates. Biogeosci. Discuss. 1–30. https://doi.org/10.5194/bg-2016-527.
Fox, D.G., 1981. Judging air quality model performance. Bull. Am. Meteorol. Soc. 62 (5), 599–609.
Gebbers, R., Adamchuk, V.I., 2010. Precision agriculture and food security. Science 327 (5967), 828–831.
Geesing, D., Diacono, M., Schmidhalter, U., 2014. Site-specific effects of variable water supply and nitrogen fertilisation on winter wheat. J. Plant Nutr. Soil Sci. 177 (4), 509–523.
Grosz, B., Dechow, R., Gebbert, S., Hoffmann, H., Zhao, G., Constantin, J., Raynal, H., Wallach, D., Coucheney, E., Lewan, E., Eckersten, H., Specka, X., Kersebaum, K.C., Nendel, C., Kuhnert, M., Yeluripati, J., Haas, E., Teixeira, E., Bindi, M., Trombi, G., Moriondo, M., Doro, L., Roggero, P.P., Zhao, Z., Wang, E., Tao, F., Rötter, R., Kassie, B., Cammarano, D., Asseng, S., Weihermüller, L., Siebert, S., Gaiser, T., Ewert, F., 2017. The implication of input data aggregation on up-scaling soil organic carbon changes. Environ. Model. Softw. 96, 361–377.
Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B.T., Kersebaum, K.C., Kleine, C., Kuhnert, M., Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P.P., Rötter, R.P., Siebert, S., Specka, X., Tao, F., Teixeira, E., Trombi, G., Wallach, D., Weihermüller, L., Yeluripati, J., Ewert, F., 2016. Impact of spatial soil and climate input data aggregation on regional yield simulations. PLoS One 11 (4), e0151782.
Hollis, J.M., Hannam, J., Bellamy, P.H., 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. Eur. J. Soil Sci. 63 (1), 96–109.
IUSS Working Group WRB, 2015. World reference Base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
Joernsgaard, B., Halmoe, S., 2003. Intra-field variation over crops and years. Eur. J. Agron. 19, 23–33.
Jürschik, P., 1999. Precision farming. Bilanz und perspektiven. DLGMitteilungen 11, 15–19.
Kersebaum, K.C., Nendel, C., 2014. Site-specific impacts of climate change on wheat production across regions of Germany using different CO2 response functions. Eur. J. Agron. 52, 22–32.
Kersebaum, K.C., Lorenz, K., Reuter, H.I., Wendroth, O., 2002. Modelling crop growth and nitrogen dynamics for advisory purposes regarding spatial variability. Agricultural System Models in Field Research and Technology Transfer. Lewis Publishers, Boca Raton, USA, pp. 229–252.
Kersebaum, K.C., Lorenz, K., Reuter, H.I., Schwarz, J., Wegehenkel, M., Wendroth, O., 2005. Operational use of agro-meteorological data and GIS to derive site specific nitrogen fertilizer recommendations based on the simulation of soil and crop growth processes. Phys. Chem. Earth 30, 59–67.
Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P.J., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. Environ. Model. Softw. 72, 402–417.
Kirkegaard, J.A., Lilley, J.M., Howe, G.N., Graham, J.M., 2007. Impact of subsoil water use on wheat yield. Aust. J. Agric. Res. 58 (4), 303–315.
Kollas, C., Kersebaum, K.C., Nendel, C., Manevski, K., Müller, C., Palosuo, T., Armas-Herrera, C.M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., Cortazar-Atauri, I.G., Giglio, L., Hlavinka, P., Hoffmann, H., Hoffmann, M.P., Launay, M., Manderscheid, R., Mary, B., Mirschel, W., Moriondo, M., Olesen, J.E., Öztürk, I., Pacholski, A., Ripoche-Wachter, D., Roggero, P.P., Roncossek, S., Rötter, R.P., Ruget, F., Sharif, B., Trnka, M., Ventrella, D., Waha, K., Wegehenkel, M., Weigel, H.-J., Wu, L., 2015. Crop rotation modelling - a European model intercomparison. Eur. J. Agron. 70, 98–111.
Kuhnert, M., Yeluripati, J., Smith, P., Hoffmann, H., Van Oijen, M., Constantin, J., Coucheney, E., Dechow, R., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K.C., Kiese, R., Klatt, S., Lewan, E., Nendel, C., Raynal, H., Sosa, C., Specka, X., Teixeira, E., Wang, E., Weihermüller, L., Zhao, G., Zhao, Z., Ogle, S., Ewert, F., 2017. Impact analysis of climate data aggregation at different spatial scales on simulated Net Primary Productivity for croplands. Eur. J. Agron. 88, 41–52.
Mann, H., Whitney, D., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18, 50–60.
Martre, P., Wallach, D., Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Rötter, R.P., Cammarano, D., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J.E., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Evaluating an ensemble of 27 crop simulation models in diverse environments: are multi-models better than one? Glob. Chang. Biol. 21 (2), 911–925. https://doi.org/10.1111/gcb.12768.
McDonald, G.K., Taylor, J.D., Verbyla, A., Kuchel, H., 2013. Assessing the importance of subsoil constraints to yield of wheat and its implications for yield improvement. Crop Pasture Sci. 63, 1043–1065.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57 (2), 443–452.
Nendel, C., Kersebaum, K.C., Mirschel, W., Wenkel, K.O., 2014. Testing farm management options as climate change adaptation strategies using the MONICA model. Eur. J. Agron. 52, 47–56.
Pachepsky, Y., Acock, B., 1998. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. Geoderma 85 (2), 213–229.
Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Şaylan, L., Šiška, B., Rötter, R.P., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35, 103–114.
Pätzold, S., Mertens, F.M., Bornemann, L., Koleczek, B., Franke, J., Feilhauer, H., Welp, G., 2008. Soil heterogeneity at the field scale: a challenge for precision crop rotation. Precis. Agric. 9, 367–390.
Rosenzweig, C., Parry, M.L., 1994. Potential impact of climate change on world food supply. Nature 367 (6459), 133–138.
Rötter, R., Van de Geijn, S.C., 1999. Climate change effects on plant growth, crop yield and livestock. Clim. Change 43 (4), 651–681.
Rötter, R.P., Palosuo, T., Kersebaum, K.-C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Olesen, J.E., Takáč, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop growth models. Field Crops Res. 133, 23–36.
Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136 (1), 235–244.
Stafford, J.V., Lark, R.M., Bolam, H.C., 1999. Using yield maps to regionalize fields into potential management units. Precision Agriculture (precisionagric4a) 225–237.
Taylor, S.L., Payton, M.E., Raun, W.R., 1999. Relationship between mean yield, coefficient of variation, mean square error, and plot size in wheat field experiments. Commun. Soil Sci. Plant Anal. 30, 1439–1447.
Trnka, M., Rötter, R.P., Ruiz-Ramos, M., Kersebaum, K.C., Olesen, J.E., Žalud, Z., Semenov, M.A., 2014. Adverse weather conditions for European wheat production will become more frequent with climate change. Nat. Clim. Change 4 (7), 637–643.
Ventrella, D., Garofalo, P., 2016. Impact of climate change on winter durum wheat

cultivated in Southern Italy: effect of extreme weather events. 8th International Congress on Environmental Modelling and Software in Toulouse. https://doi.org/10.13140/RG.2.2.21794.63680. Poster 252.

Ventrella, D., Stellacci, A.M., Castrignanò, A., Charfeddine, M., Castellini, M., 2016. Effects of crop residue management on winter durum wheat productivity in a long term experiment in Southern Italy. Eur. J. Agron. 77, 188–198.

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37 (1), 29–38.

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P.J., van Ittersum, M., Aggarwal, P.K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Dumont, B., Rezaei, E.E., Fereres, E., Fitzgerald, G.J., Gao, Y., Garcia-Vila, M., Gayler, S., Girousse, C., Hoogenboom, G., Horan, H., Izaurralde, R.C., Jones, C.D., Kassie, B.T., Kersebaum, K.C., Klein, C., Koehler, A.K., Maiorano, A., Minoli, S., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G.J., Palosuo, T., Priesack, E., Ripoche, D., Rötter, R.P., Semenov, M.A., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Wolf, J., Zhang, Z., 2018. Multi-model ensembles improve predictions of crop-environment-management interactions. Glob. Change Biol. https://doi.org/10.1111/gcb.14411. in press.

Wallor, E., Kersebaum, K.C., Lorenz, K., Gebbers, R., 2018. A comprehensive dataset demonstrating the within-field variability of soil properties and crop growth conditions in northwestern Germany. Open Data J. Agric. Res (in press).

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Liu, L., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Rezaei, E.E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., Asseng, S., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nat. Plants 3, 17102. https://doi.org/10.1038/nplants.2017.102.

Wassenaar, T., Lagacherie, P., Legros, J.P., Rounsevell, M.D.A., 1999. Modelling wheat yield responses to soil and climate variability at the regional scale. Clim. Res. 11 (3), 209–220.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Wong, M.T.F., Asseng, S., 2006. Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model. Plant Soil 283 (1-2), 203–215.

Wu, L., Kersebaum, K.C., 2008. Modeling water and nitrogen interaction responses and their consequences in crop models. In: In: Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Q. (Eds.), Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects on Plant Growth Processes. Advances in Agricultural Systems Modeling, vol. 1. ASA, CSSA, SSSA, Madison, WI, USA, pp. 215–249.

Yin, X., Kersebaum, K.C., Kollas, C., Baby, S., Beaudoin, N., Manevski, K., Palosuo, T., Nendel, C., Wu, L., Hoffmann, H., 2017. Multi-model uncertainty analysis in predicting grain N for crop rotations in Europe. Eur. J. Agron. 84, 152–165.

Zambrano-Bigiarini, M., 2014. hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. R package version 0.3-8. http://CRAN.R-project.org/package=hydroGOF.

Zhao, G., Hoffmann, H., Yeluripati, J., Specka, X., Nendel, C., Coucheney, E., Kuhnert, M., Tao, F., Constantin, J., Raynal, H., Teixeira, E., Grosz, B., Doro, L., Kiese, R., Eckersten, H., Haas, E., Cammarano, D., Kassie, B., Moriondo, M., Trombi, G., Bindi, M., Biernath, C., Heinlein, F., Klein, C., Priesack, E., Lewan, E., Kersebaum, K.C., Rötter, R.P., Roggero, P.P., Wallach, D., Asseng, S., Siebert, S., Gaiser, T., Ewert, F., 2016. Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. Environ. Model. Softw. 80, 100–112.